Further Topics on Random Variables

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Probability

Outline

- Derived Distribution
- Covariance and Correlation
- Conditional Expectation
- Moment Generating Function

2-step method

Let X be CRV. Consider $Y = g(X)$ where $g(\cdot)$ is differentiable. The distribution of Y can be derived.

- Equivalent events in terms of X and Y

$$(Y \leq y) = (g(X) \leq y) = \bigcup_i (l_i(y) \leq X \leq u_i(y))$$

$$\Rightarrow \; P(Y \leq y) = \sum_i P(X \leq u_i(y)) - P(X \leq l_i(y))$$

- Find CDF and PDF

$$F_Y(y) = \sum_i F_X(u_i(y)) - F_X(l_i(y))$$

$$\Rightarrow \; f_Y(y) = \frac{d}{dy} F_Y(y) = \sum_i f_X(u_i(y))u_i'(y) - f_X(l_i(y))l_i'(y)$$

Example (4.1 2-step method)

Let X be **Uni**$(0,1)$ and $Y = \sqrt{X}$. Derive f_Y.

The range of Y is $\mathcal{Y} = (0,1)$. For $y \in \mathcal{Y}$, we have equivalent events $(Y \le y) = \left(\sqrt{X} \le y\right) = (X \le y^2)$, and

$$P(Y \le y) = P\left(X \le y^2\right) = F_X(y^2) = y^2$$

Thus

$$F_Y(y) = \begin{cases} 0, & y \le 0 \\ y^2, & 0 < y < 1 \\ 1, & y \ge 1 \end{cases}$$

$$\Rightarrow \ f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Example (4.2 2-step method)

Scott is biking from Seattle to Portland, a distance of 180 miles, at a constant speed which is uniformly distributed between 30 and 60 miles per hour. Derive the PDF of the duration of the trip.

Let $X \sim \textbf{Uni}(30, 60)$ and $Y = \frac{180}{X}$ be duration. For $y \in \mathcal{Y} = (3, 6)$

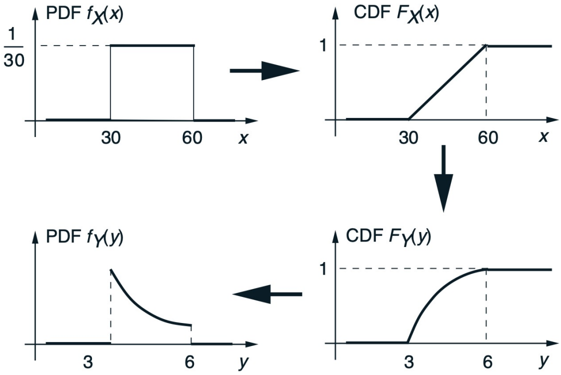
$$(Y \le y) = \left(\frac{180}{X} \le y\right) = \left(X \ge \frac{180}{y}\right) = \left(X < \frac{180}{y}\right)^c$$

$$\Rightarrow \; P(Y \le y) = 1 - P\left(X < \frac{180}{y}\right) = 1 - F_X\left(\frac{180}{y}\right)$$

$$= 1 - \frac{1}{30}\left(\frac{180}{y} - 30\right) = 2 - \frac{6}{y}$$

Thus

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} \frac{6}{y^2}, & 3 < y < 6 \\ 0, & \text{otherwise} \end{cases}$$


PDF $f_X(x)$
$\frac{1}{30}$
30
60
x
CDF $F_X(x)$
1
30
60
x
CDF $F_Y(y)$
1
3
6
y
PDF $f_Y(y)$
3
6
y

Example (4.3 2-step method)

Let X be a CRV and $Y = X^2$. Derive f_Y from f_X.

For $y \leq 0$, we have $P(Y \leq y) = 0$. For $y > 0$

$$(Y \leq y) = \left(X^2 \leq y\right) = (|X| \leq \sqrt{y}) = (-\sqrt{y} \leq X \leq \sqrt{y})$$

$$P(Y \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = P(X \leq \sqrt{y}) - P(X < -\sqrt{y})$$

$$\Rightarrow \; F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Thus

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}), & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Linear function

Let X be a CRV and $Y = aX + b$. We have

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Case: $a > 0$

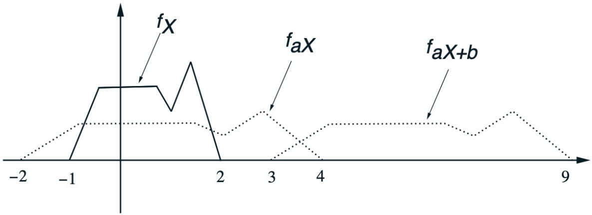
$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right)$$

$$\Rightarrow \ F_Y(y) = F_X\left(\frac{y-b}{a}\right) \ \Rightarrow \ f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Case: $a < 0$

$$P(Y \leq y) = P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right)$$

$$\Rightarrow \ F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right) \ \Rightarrow \ f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$



$$a = 2 \text{ and } b = 5$$

Example (Scaled exponential)

Let X be **Exp**(λ) and $Y = aX$ where $a > 0$. Then

$$Y \sim \textbf{Exp}\left(\frac{\lambda}{a}\right)$$

The PDF of $Y = aX + b$ is related to the PDF of X by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Here $b = 0$ and $f_X(x) = \lambda e^{-\lambda x} u(x)$, so

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y}{a}\right) = \frac{1}{a}\lambda e^{-\lambda(\frac{y}{a})} u\left(\frac{y}{a}\right) = \frac{\lambda}{a} e^{-(\frac{\lambda}{a})y} u(y)$$

That is

$$Y \sim \textbf{Exp}\left(\frac{\lambda}{a}\right)$$

Example (Linear function of Normal)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = aX + b$. The PDF of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}} \end{aligned}$$

That is

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

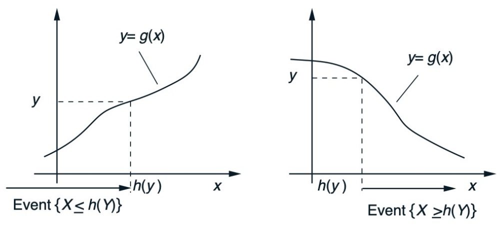
Monotone function

Let X be CRV. Consider $Y = g(X)$ where $g(\cdot)$ is strictly monotone and differentiable over range $\mathcal{X}$.

- From $\mathcal{X}$ and $g(\cdot)$, we can decide range $\mathcal{Y}$
- $\exists\, h(\cdot)$ such that $y = g(x)$ iff $x = h(y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$
- For $y \in \mathcal{Y}$, the PDF of Y is

$$f_Y(y) = f_X(h(y))|h'(y)|$$

CDF argument



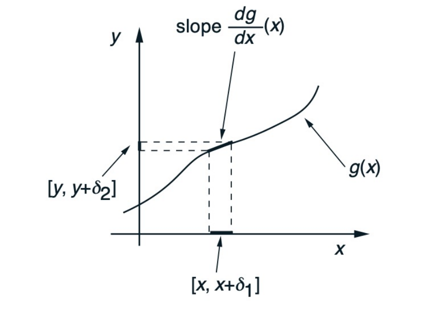
Case: $g(x)$ increasing

$$(Y \le y) = (X \le h(y))$$
$$\Rightarrow F_Y(y) = \int_{-\infty}^{h(y)} f_X(x)dx$$
$$\Rightarrow f_Y(y) = f_X(h(y))h'(y)$$

Case: $g(x)$ decreasing

$$(Y \le y) = (X \ge h(y)) = (X < h(y))^c$$
$$\Rightarrow F_Y(y) = 1 - \int_{-\infty}^{h(y)} f_X(x)dx$$
$$\Rightarrow f_Y(y) = -f_X(h(y))h'(y)$$

PDF argument



$$
\begin{aligned}
(Y \in [y, y+\delta_2]) = (X \in [x, x+\delta_1]) &\Rightarrow f_Y(y)\delta_2 + o(\delta_2) = f_X(x)\delta_1 + o(\delta_1) \\
&\Rightarrow \frac{f_Y(y)}{f_X(x)} = \frac{\delta_1}{\delta_2} = \left(\frac{\delta_2}{\delta_1}\right)^{-1} = \left(\frac{dg(x)}{dx}\right)^{-1} \\
&\Rightarrow f_Y(y) = f_X(x)\left(g'(x)\right)^{-1} = f_X(h(y))\left(h'(y)\right)
\end{aligned}
$$

Example (4.2 again)

Duration $Y = \frac{180}{X} = g(X)$ is monotone over $\mathcal{X} = (30, 60)$ with inverse $h(y) = \frac{180}{y}$. For $y \in (3, 6)$, we have

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \left| h'(y) \right| \\ &= \frac{1}{30}(u(h(y) - 30) - u(h(y) - 60)) \left| \left(\frac{180}{y} \right)' \right| \\ &= \frac{1}{30} \left(u \left(\frac{180}{y} - 30 \right) - u \left(\frac{180}{y} - 60 \right) \right) \frac{180}{y^2} \\ &= \frac{6}{y^2} \left(u \left(6 - y \right) - u \left(3 - y \right) \right) \\ &= \frac{6}{y^2} \left(u \left(y - 3 \right) - u \left(y - 6 \right) \right) \end{aligned}$$

This is consistent with what we found earlier by the 2-step method.

Example (4.6)

Let X be **Uni**$(0,1)$ and $Y = X^2$. Derive f_Y.

Function $Y = X^2 = g(X)$ is monotone over $\mathcal{X} = (0,1)$ with inverse $h(y) = \sqrt{y}$. For $y \in (0,1)$, we have

$$\begin{aligned} f_Y(y) &= f_X(h(y)) \left| h'(y) \right| \\ &= \left(u\left(\sqrt{y} - 0\right) - u\left(\sqrt{y} - 1\right)\right) \frac{1}{2\sqrt{y}} \\ &= \frac{1}{2\sqrt{y}} \end{aligned}$$

Multi-variate function

Let X and Y be CRVs. Consider $Z = g(X, Y)$ where $g(\cdot, \cdot)$ is differentiable. The distribution of Z can be derived.

The CDF of Z is

$$P(Z \leq z) = P(g(X, Y) \leq z) = P((X, Y) \in \{(x, y) \,|\, g(x, y) \leq z\})$$

Defining region $R_z = \{(x, y) \,|\, g(x, y) \leq z\}$, we have

$$P(Z \leq z) = P((X, Y) \in R_z) = \iint_{R_z} f_{XY}(x, y)\, dxdy$$

R_z depends on z. It is on one side of $g(x, y) = z$.

Example (4.7 2-step method)

The distance of a shot off the center of a target is **Uni**$(0,1)$. For 2 shots, derive the PDF of the distance of the losing shot.

Let X and Y be the shot distances. Then $Z = \max(X,Y)$ is the distance of the losing shot. For $z \in [0,1]$, we have

$$R_z = \{(x,y) \mid \max(x,y) \le z\}$$

R_z is a square with side z, so

$$P(Z \le z) = \iint_{R_z} f_{XY}(x,y)\, dxdy = z^2$$

Thus

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} 2z, & 0 \le z \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Example (4.8 2-step method)

Let X and Y be independent **Uni**$(0,1)$. Derive the PDF of $Z = \frac{Y}{X}$.

For $z \leq 0$, we have $P(Z \leq z) = 0$.

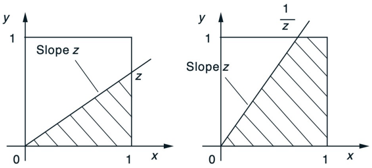
For $z > 0$, we have

$$R_z = \left\{ (x,y) \mid \left(\frac{y}{x} \right) \leq z \right\} = \{(x,y) \mid y \leq zx\}$$

Note R_z is on the right side of the line $y = zx$. With R_z, we have

$$F_Z(z) = P(Z \leq z) = \iint_{R_z} f_{XY}(x,y)\, dxdy$$

There are 2 cases to consider (next page).



Case: $0 < z < 1$

$$P(Z \leq z) = \int_0^1 \int_0^{zx} 1 \, dydx = \int_0^1 zx \, dx = \frac{z}{2}$$

Case: $z > 1$

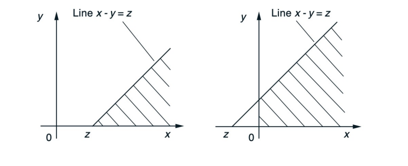
$$P(Z \leq z) = \int_0^1 \int_{\frac{y}{z}}^1 1 \, dxdy = \int_0^1 \left(1 - \frac{y}{z}\right) dy = 1 - \frac{1}{2z}$$

Combining both cases, we have

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \begin{cases} \frac{1}{2}, & 0 < z < 1 \\ \frac{1}{2z^2}, & z > 1 \\ 0, & \text{otherwise} \end{cases}$$

Example (4.9 2-step method)

Let X and Y be independent exponential random variables with the same parameter λ, and $Z = X - Y$. Derive the PDF of Z.



For any z, we have

$$R_z = \{(x, y) \mid x - y \leq z\}$$

Note R_z is on the left side of the line $x - y = z$ (opposite the shaded region). With R_z, we have

$$F_Z(z) = P(Z \leq z) = \iint_{R_z} f_{XY}(x, y)\, dxdy$$

There are 2 cases to consider (next page).

Case: $z \geq 0$

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z+y} f_{XY}(x, y)\, dx\, dy$$

$$= \int_0^{\infty} \lambda e^{-\lambda y} \int_0^{z+y} \lambda e^{-\lambda x} dx dy = \int_0^{\infty} \lambda e^{-\lambda y} \left(1 - e^{-\lambda(z+y)}\right) dy$$

$$= 1 - e^{-\lambda z} \int_0^{\infty} \lambda e^{-\lambda(2y)} dy = 1 - \frac{1}{2} e^{-\lambda z}$$

Case: $z < 0$

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{x-z}^{\infty} f_{XY}(x, y) dy dx$$

$$= \int_0^{\infty} \lambda e^{-\lambda x} \int_{x-z}^{\infty} \lambda e^{-\lambda y}\, dy\, dx = \int_0^{\infty} \lambda e^{-\lambda x} e^{-\lambda(x-z)} dx$$

$$= e^{\lambda z} \int_0^{\infty} \lambda e^{-\lambda(2x)} dx = \frac{1}{2} e^{\lambda z}$$

For both cases, we have

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{\lambda}{2} e^{-\lambda|z|}$$

Sum

- Let X and Y be DRVs and $Z = X + Y$. The PMF of Z is

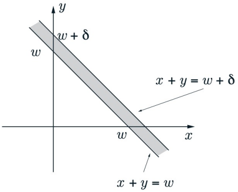
$$p_Z(z) = \sum_{x \in \mathcal{X}} p_X(x) p_{Y|X}(z - x|x)$$

- Let X and Y be CRVs and $W = X + Y$. The PDF of W is

$$f_W(w) = \int f_X(x) f_{Y|X}(w - x|x)\, dx$$

Case: DRV

$$
\begin{aligned}
(Z = z) = (X + Y = z) &= \left(\bigcup_{x \in \mathcal{X}} (X = x) \right) \cap (X + Y = z) \\
&= \bigcup_{x \in \mathcal{X}} ((X = x) \cap (X + Y = z)) \\
\Rightarrow \quad p_Z(z) &= P \left(\bigcup_{x \in \mathcal{X}} ((X = x) \cap (X + Y = z)) \right) \\
&= \sum_{x \in \mathcal{X}} P\left((X = x) \cap (X + Y = z)\right) \\
&= \sum_{x \in \mathcal{X}} P((X = x) \cap (Y = z - X)) \\
&= \sum_{x \in \mathcal{X}} P(X = x) P(Y = z - X | X = x) \\
&= \sum_{x \in \mathcal{X}} p_X(x) p_{Y|X}(z - x | x)
\end{aligned}
$$



Case: CRV. Note $P(W \in (w, w+\delta)) = P((X,Y)$ in the shaded strip).

$$f_W(w)\delta = \int_{-\infty}^{\infty} \int_{w-x}^{w+\delta-x} f_{XY}(x,y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} \int_{w-x}^{w-x+\delta} f_X(x) f_{Y|X}(y|x)\, dy\, dx$$

$$= \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(w-x|x)\, \delta\, dx$$

$$\Rightarrow \ f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(w-x|x)\, dx$$

Independent sum

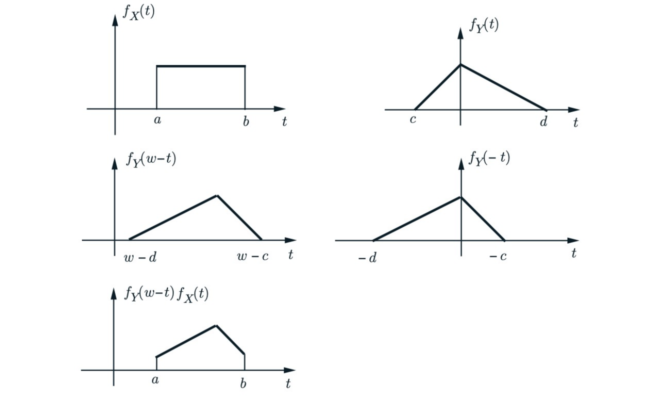
- Let X and Y be independent DRVs and $Z = X + Y$. The PMF of Z is

$$p_Z(z) = \sum_{x \in \mathcal{X}} p_X(x) p_Y(z - x)$$

- Let X and Y be independent CRVs and $W = X + Y$. The PDF of W is

$$f_W(w) = \int f_X(t) f_Y(w - t) dt$$

Convolution diagram



Example (4.10 Uniform convolution)

Let X, Y be independent **Uni**$(0,1)$. Find the PDF of $Z = X + Y$.

$$\begin{aligned} f_Z(z) &= \int f_X(x) f_Y(z-x)\,dx \\ &= \int_0^1 (u(z-x) - u((z-x)-1))\,dx \\ &= \int_0^1 (u(x-(z-1)) - u(x-z))\,dx \\ &= \begin{cases} z, & 0 < z \leq 1 \\ 2-z, & 1 < z < 2 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

Note $f_Z(z)$ can be found via $R_z = \{(x,y) \mid x + y \leq z\}$.

Example (4.11 Normal convolution)

Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be independent. Find the PDF of $Z = X + Y$.

$$f_Z(z) = \int f_X(x) f_Y(z-x)\, dx$$

$$= \int \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} e^{-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}}\, dx$$

It can be shown that $f_Z(z)$ is exponential with a (negative) quadratic exponent, so Z is Normal. Since

$$\mathbf{E}[Z] = \mathbf{E}[X] + \mathbf{E}[Y] = \mu_X + \mu_Y$$

$$\textbf{var}(Z) = \textbf{var}(X) + \textbf{var}(Y) = \sigma_X^2 + \sigma_Y^2$$

we have

$$Z \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Example (4.12)

Let X and Y be independent CRVs. The PDF of $Z = X - Y$ is

$$f_Z(z) = \int f_X(x) f_{-Y}(z-x)\,dx = \int f_X(x) f_Y(x-z)\,dx$$

Consider independent $X \sim$ **Exp**(λ) and $Y \sim$ **Exp**(λ), and $Z = X - Y$.

$$\begin{aligned} f_Z(z) &= \int f_X(x) f_Y(x-z)\,dx \\ &= \int_0^\infty \lambda e^{-\lambda x} \lambda e^{-\lambda(x-z)} u(x-z)\,dx \\ &= \lambda^2 e^{\lambda z} \int_{\max(0,z)}^\infty e^{-2\lambda x}\,dx \\ &= \frac{\lambda}{2} e^{-\lambda|z|} \end{aligned}$$

Definition (covariance and correlation)

Let X and Y be random variables.

- The covariance of X and Y is defined by

$$\textbf{cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

- The correlation coefficient of X and Y is defined by

$$\textbf{Corr}(X,Y) = \frac{\textbf{cov}(X,Y)}{\sqrt{\textbf{var}(X)}\sqrt{\textbf{var}(Y)}}$$

  $\textbf{Corr}(X,Y)$ is also denoted by ρ_{XY}.

Properties

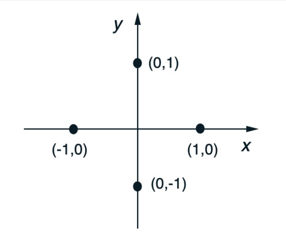
Let X and Y be random variables.

$$\textbf{cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$$
$$\textbf{cov}(X,Y) = \textbf{cov}(Y,X)$$
$$\textbf{cov}(X,aY+b) = a\ \textbf{cov}(X,Y)$$
$$\textbf{cov}(X,Y+Z) = \textbf{cov}(X,Y) + \textbf{cov}(X,Z)$$

$$-1 \leq \textbf{Corr}(X,Y) \leq 1$$

Example (4.13)

Let X and Y be DRVs. The joint PMF of X and Y has value $\frac{1}{4}$ on points $(1, 0), (0, 1), (-1, 0), (0, -1)$. Compute **cov**$(X, Y)$ and **Corr**(X, Y).



Sum variance

Let X and Y be random variables. Then

$$\textbf{var}(X+Y) = \textbf{var}(X) + \textbf{var}(Y) + 2\,\textbf{cov}(X,Y)$$

More generally, let $X_1, \ldots, X_n$ be random variables. Then

$$\textbf{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \textbf{var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \textbf{cov}(X_i, X_j)$$

Example (4.15)

There are n persons, each with a hat. They put their hats in a box and then each retrieves a hat. Let H be the number of persons who retrieve own hats. Find the variance of H.

Let X_i be the indicator for person i retrieving own hat. Note $H = X_1 + \cdots + X_n$, so **var**(H) consists of terms of the variances and covariances of X_i's.

- To find the variance of X_i, we need $\mathbf{E}[X_i]$ and $\mathbf{E}[X_i^2]$, which depends on $P(X_i = 1)$
- To find the covariance of X_i and X_j, we need $\mathbf{E}[X_i X_j]$, which depends on $P(X_i = 1 \cap X_j = 1)$
- We can find $P(X_i = 1)$ and $P(X_i = 1 \cap X_j = 1)$ through I_i, the event that hat i is in the box as person i is retrieving

$$\begin{aligned}
P(X_i = 1) &= P(X_i = 1 \cap I_i) + \overbrace{P(X_i = 1 \cap I_i^c)}^{0} \\
&= P(I_i) P(X_i = 1 | I_i) \\
&= \left(\frac{n-i+1}{n} \right) \left(\frac{1}{n-i+1} \right) \\
&= \frac{1}{n}
\end{aligned}$$

$$\begin{aligned}
P(X_i = 1 \cap X_j = 1) &= P(X_i = 1) P(X_j = 1 | X_i = 1) \\
&= P(X_i = 1)(P(X_j = 1 \cap I_j | X_i = 1) + \underbrace{P(X_j = 1 \cap I_j^c | X_i = 1)}_{0}) \\
&= P(X_i = 1) P(I_j | X_i = 1) P(X_j = 1 | I_j \cap X_i = 1) \\
&= \frac{1}{n} \frac{n-j+1}{n-1} \frac{1}{n-j+1} \\
&= \frac{1}{n} \cdot \frac{1}{n-1}
\end{aligned}$$

$$\begin{aligned}
\textsf{var}(X_i) &= \mathbf{E}\left[X_i^2\right] - \mathbf{E}^2\left[X_i\right] \\
&= \frac{1}{n} - \frac{1}{n^2} \\
&= \frac{n-1}{n^2} \\
\textsf{cov}(X_i, X_j) &= \mathbf{E}[X_i X_j] - \mathbf{E}[X_i]\mathbf{E}[X_j] \\
&= P(X_i = 1 \cap X_j = 1) - \mathbf{E}[X_i]\mathbf{E}[X_j] \\
&= \frac{1}{n} \cdot \frac{1}{n-1} - \left(\frac{1}{n}\right)^2 \\
&= \frac{1}{n^2(n-1)}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \ \textsf{var}(H) &= \sum_{i=1}^{n} \textsf{var}(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \textsf{cov}(X_i, X_j) \\
&= n\left(\frac{n-1}{n^2}\right) + 2\,\frac{n(n-1)}{2}\,\frac{1}{n^2(n-1)} \\
&= \frac{n-1}{n} + \frac{1}{n} \\
&= 1
\end{aligned}$$

Definition (conditional expectation)

Let X and Y be CRVs.

- The expectation of X conditional on $Y = y$ is

$$\mathbf{E}[X|Y = y] = \int x \, f_{X|Y}(x|y) \, dx$$

- The conditional expectation of X given Y is defined by

$$\mathbf{E}[X|Y] = g(Y) \text{ where } g(y) = \mathbf{E}[X|Y = y]$$

Expectation and variance

Consider the conditional expectation $\mathbf{E}[X|Y] = g(Y)$.

- As a function of Y, $\mathbf{E}[X|Y]$ is a random variable
- The expectation of $\mathbf{E}[X|Y]$ is

$$\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[g(Y)]$$

- The variance of $\mathbf{E}[X|Y]$ is

$$\textsf{var}(\mathbf{E}[X|Y]) = \textsf{var}(g(Y)) = \mathbf{E}\left[(g(Y) - \mathbf{E}[g(Y)])^2\right]$$

Example (4.16 Conditional expectation)

The probability of heads of a coin is uncertain, so we may assume it is drawn from the distribution of a random variable Y. Let X be the number of heads in n tosses of this coin. Find the conditional expectation of X given Y.

The expectation of X conditional on $Y = y$ is

$$\mathbf{E}[X|Y = y] = ny = g(y)$$

So the conditional expectation of X given Y is

$$\mathbf{E}[X|Y] = g(Y) = nY$$

Law of iterated expectation

Let X and Y be CRVs. Then

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$$

In order to find $\mathbf{E}[X]$, we can incorporate any random variable Y, find $E[X|Y]$ first, then find its expectation.

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}[g(Y)] = \int g(y)\, f_Y(y)\, dy &= \int \mathbf{E}[X|Y=y]\, f_Y(y)\, dy \\
&= \int \left(\int x\, f_{X|Y}(x|y) dx \right) f_Y(y)\, dy = \iint x\, f_{XY}(x,y)\, dx\, dy \\
&= \int x\, f_X(x)\, dx = \mathbf{E}[X]
\end{aligned}$$

Scenarios for iterated expectation

- X is the random variable of interest
- Something crucial about the distribution of X is unknown
- Had it been known, the distribution of X would be simple
- Introduce random variable(s) for the unknown

Consider the previous example again.

- X is the number of heads
- As the probability of heads is unknown, the distribution of X is not clear
- Introduce random variable Y for the probability of heads
- Conditional on $Y = y$, X is binomial

Example (4.16 Iterated expectation)

For $Y \sim \textbf{Uni}(0,1)$, find $\mathbf{E}[X]$.

By the law of iterated expectation

$$\mathbf{E}[X] = \mathbf{E}\left[\mathbf{E}[X|Y]\right] = \mathbf{E}[nY] = n\mathbf{E}[Y] = \frac{n}{2}$$

Alternatively, $\mathbf{E}[X]$ can be computed from the PMF of X

$$p_X(x) = \int p_{X|Y}(x|y) f_Y(y) dy = \int_0^1 \binom{n}{x} y^x (1-y)^{n-x} (1) dy = \frac{1}{n+1}$$

$$\Rightarrow \quad \mathbf{E}[X] = \sum_{x=0}^{n} x\, p_X(x) = \frac{n}{2}$$

Example (4.17 Conditional expectation)

We start with a stick with length l. We break it at a random point uniformly distributed over the stick and keep the left part. The same process is repeated on the remaining part. Compute the expectation of the length of the final part after breaking **twice**.

Let Y be the length after one break and X be the length after the second break. By the law of iterated expectation

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]] = \mathbf{E}\left[\frac{Y}{2}\right] = \frac{1}{2}\mathbf{E}\left[Y\right] = \frac{l}{4}$$

Alternatively, $\mathbf{E}[X]$ can be computed from the PDF of X with range $\mathcal{X} = [0, l]$. For $x \in \mathcal{X}$, we have

$$
\begin{aligned}
f_X(x) &= \int f_{X|Y}(x|y)\, f_Y(y)\, dy \\
&= \frac{1}{l} \int_0^l \frac{1}{y}(u(x) - u(x-y))dy \\
&= \frac{1}{l} \left(\int_0^x \frac{1}{y} \cdot 0\, dy + \int_x^l \frac{1}{y} \cdot 1\, dy \right) \\
&= \frac{1}{l} \log \frac{l}{x} \\
\Rightarrow \quad \mathbf{E}[X] &= \int x f_X(x)\, dx = \int_0^l x \left(\frac{1}{l} \log \frac{l}{x} \right) dx \\
&= \frac{1}{l} \left(\int_0^l x \log l\, dx - \int_0^l x \log x\, dx \right) \\
&= \frac{1}{l} \left(\log l \frac{l^2}{2} - \left[\frac{x^2}{2} \log x \Big|_0^l - \int_0^l \frac{x^2}{2} (\log x)'\, dx \right] \right) \\
&= \frac{l}{4}
\end{aligned}
$$

Definition (conditional variance)

Let X and Y be CRVs.

- The variance of X conditional on $Y = y$ is

$$\textbf{var}(X|Y=y) = \mathbf{E}\left[(X - \mathbf{E}[X|Y=y])^2 \,|\, Y=y\right]$$

- The conditional variance of X given Y is defined by

$$\textbf{var}(X|Y) = h(Y) \text{ where } h(y) = \textbf{var}(X|Y=y)$$

Conditional variance formula

$$\mathsf{var}(X|Y) = \mathbf{E}\left[X^2|Y\right] - \mathbf{E}^2[X|Y]$$

$$\begin{aligned}
\mathsf{var}(X|Y=y) &= \mathbf{E}\left[(X - \mathbf{E}[X|Y=y])^2 \mid Y=y\right] \\
&= \int (x - \mathbf{E}[X|Y=y])^2 f_{X|Y}(x|y)\,dx \\
&= \int (x^2 - 2x\mathbf{E}[X|Y=y] + \mathbf{E}^2[X|Y=y])\, f_{X|Y}(x|y)\,dx \\
&= \int x^2 f_{X|Y}(x|y)\,dx - 2\mathbf{E}^2[X|Y=y] + \mathbf{E}^2[X|Y=y] \\
&= \mathbf{E}\left[X^2|Y=y\right] - \mathbf{E}^2[X|Y=y] \\
&= f(y) - g(y)
\end{aligned}$$

$$\Rightarrow\ \mathsf{var}(X|Y) = f(Y) - g(Y) = \mathbf{E}\left[X^2|Y\right] - \mathbf{E}^2[X|Y]$$

Expectation and variance

Let X and Y be CRVs. Consider **var**$(X|Y) = h(Y)$.

- As a random variable, **var**$(X|Y)$ has expectation and variance
- It has been shown that

$$\mathbf{var}(X|Y) = \mathbf{E}\left[X^2|Y\right] - \mathbf{E}^2\left[X|Y\right]$$

so the expectation of **var**$(X|Y)$ is

$$\mathbf{E}[\mathbf{var}(X|Y)] = \mathbf{E}\left[\mathbf{E}\left[X^2|Y\right]\right] - \mathbf{E}\left[\mathbf{E}^2\left[X|Y\right]\right]$$

- The variance of **var**$(X|Y)$ is

$$\mathbf{var}(\mathbf{var}(X|Y)) = \mathbf{E}\left[(\mathbf{var}(X|Y) - \mathbf{E}[\mathbf{var}(X|Y)])^2\right]$$

Law of total variance

Let X and Y be random variables. Then

$$\mathbf{var}(X) = \mathbf{var}(\mathbf{E}[X|Y]) + \mathbf{E}[\mathbf{var}(X|Y)]$$

In order to find $\mathbf{var}(X)$, we can incorporate any random variable Y. Then $\mathbf{var}(X)$ consists of the expectation of conditional variance and the variance of conditional expectation.

$$\begin{aligned}
\mathbf{var}(\mathbf{E}[X|Y]) &= \mathbf{E}\left[(\mathbf{E}[X|Y] - \mathbf{E}[\mathbf{E}[X|Y]])^2\right] \\
&= \mathbf{E}\left[(\mathbf{E}[X|Y] - \mathbf{E}[X])^2\right] \\
&= \mathbf{E}\left[\mathbf{E}^2[X|Y] - 2\mathbf{E}[X]\mathbf{E}[X|Y] + \mathbf{E}^2[X]\right] \\
&= \mathbf{E}\left[\mathbf{E}^2[X|Y]\right] - \mathbf{E}^2[X] \\
\mathbf{E}[\mathbf{var}(X|Y)] &= \mathbf{E}\left[\mathbf{E}\left[X^2|Y\right] - \mathbf{E}^2[X|Y]\right] \\
&= \mathbf{E}\left[X^2\right] - \mathbf{E}\left[\mathbf{E}^2[X|Y]\right]
\end{aligned}$$

Example (4.16 Total variance)

For $Y \sim \textbf{Uni}(0,1)$, find $\textbf{var}(X)$.

$$\begin{aligned}
\textbf{var}(X) &= \textbf{var}(\mathbf{E}[X|Y]) + \mathbf{E}[\textbf{var}(X|Y)] \\
&= \textbf{var}(nY) + \mathbf{E}[nY(1-Y)] \\
&= n^2\textbf{var}(Y) + n\mathbf{E}[Y(1-Y)] \\
&= n^2\left(\mathbf{E}\left[Y^2\right] - \mathbf{E}^2[Y]\right) + n\left(\mathbf{E}[Y] - \mathbf{E}\left[Y^2\right]\right) \\
&= n^2\left(\frac{1}{3} - \frac{1}{4}\right) + n\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{n^2}{12} + \frac{n}{6}
\end{aligned}$$

Alternatively, using $p_X(x) = \frac{1}{n+1}$, we get

$$\textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] = \sum_{x=0}^{n} x^2 p_X(x) - \left(\frac{n}{2}\right)^2 = \frac{n^2}{12} + \frac{n}{6}$$

Example (4.17 Total variance)

Let Y be the length after one break and X be the length after the second break. Find **var**(X).

By the law of total variance

$$\textbf{var}(X) = \textbf{var}(\mathbf{E}[X|Y]) + \mathbf{E}[\textbf{var}(X|Y)]$$
$$= \textbf{var}\left(\frac{Y}{2}\right) + \mathbf{E}\left[\frac{Y^2}{12}\right]$$
$$= \frac{1}{4}\frac{l^2}{12} + \frac{1}{12}\frac{l^2}{3} = \frac{7}{144}l^2$$

Alternatively, using $f_X(x) = \frac{1}{l}\log\frac{l}{x}(u(x) - u(x-l))$, we get

$$\textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] = \int x^2 f_X(x)dx - \left(\frac{l}{4}\right)^2$$
$$= \frac{l^2}{9} - \left(\frac{l}{4}\right)^2 = \frac{7}{144}l^2$$

Definition (moment generating function)

Let X be a random variable. The moment generating function (MGF) of X is defined by

$$M_X(s) = \mathbf{E}\left[e^{sX}\right]$$

- If X is DRV, $M_X(s)$ is the transform of PMF $p_X(x)$ by

$$M_X(s) = \sum_{x \in \mathcal{X}} e^{sx} p_X(x)$$

- If X is CRV, $M_X(s)$ is the transform of PDF $f_X(x)$ by

$$M_X(s) = \int e^{sx} f_X(x) dx$$

Example (4.22 MGF of DRV)

Let X be DRV with PMF

$$p_X(x) = \begin{cases} \frac{1}{2}, & x = 2 \\ \frac{1}{6}, & x = 3 \\ \frac{1}{3}, & x = 5 \\ 0, & \text{otherwise} \end{cases}$$

Find $M_X(s)$.

$$M_X(s) = \mathbf{E}\left[e^{sX}\right] = \sum_{x \in \mathcal{X}} e^{sx} p_X(x)$$

$$= \frac{1}{2}e^{2s} + \frac{1}{6}e^{3s} + \frac{1}{3}e^{5s}$$

Example (4.23 Poisson MGF)

For $X \sim \textbf{Poi}(\lambda)$, find $M_X(s)$.

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \;\Rightarrow\; M_X(s) = \mathbf{E}\left[e^{sX}\right] = \sum_{x=0}^{\infty} e^{sx} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^s)^x}{x!} = e^{-\lambda}e^{\lambda e^s} = e^{\lambda(e^s-1)}$$

Example (4.24 Exponential MGF)

For $X \sim \textbf{Exp}(\lambda)$, find $M_X(s)$.

$$\begin{aligned} f_X(x) = \lambda e^{-\lambda x} u(x) \ \Rightarrow \ M_X(s) &= \mathbf{E}\left[e^{sX}\right] \\ &= \int_0^\infty e^{sx} \lambda e^{-\lambda x} dx \\ &= \int_0^\infty \lambda e^{-(\lambda - s)x} dx \\ &= \lambda \, \frac{-1}{\lambda - s} \, e^{-(\lambda - s)x} \Big|_0^\infty \\ &= \frac{\lambda}{\lambda - s} \end{aligned}$$

Example (4.25 Linear function MGF)

Let X be random variable with MGF $M_X(s)$ and $Y = aX + b$. Relate $M_Y(s)$ to $M_X(s)$.

$$\begin{aligned} M_Y(s) &= \mathbf{E}\left[e^{sY}\right] \\ &= \mathbf{E}\left[e^{s(aX+b)}\right] \\ &= e^{bs}\mathbf{E}\left[e^{asX}\right] \\ &= e^{bs}M_X(as) \end{aligned}$$

Example (4.26 Normal MGF)

For $X \sim \mathcal{N}(\mu, \sigma^2)$, find $M_X(s)$.

Let Z be $\mathcal{N}(0,1)$ so $X = \sigma Z + \mu$. We work on Z.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Rightarrow M_Z(s) = \mathbf{E}\left[e^{sZ}\right] = \frac{1}{\sqrt{2\pi}} \int e^{sz} e^{-\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int e^{-\frac{(z-s)^2}{2}} e^{\frac{s^2}{2}} dz$$

$$= e^{\frac{s^2}{2}} \left(\frac{1}{\sqrt{2\pi}} \int e^{-\frac{(z-s)^2}{2}} dz \right)$$

$$= e^{\frac{s^2}{2}}$$

So

$$M_X(s) = e^{\mu s} M_Z(\sigma s) = e^{\frac{1}{2}\sigma^2 s^2 + \mu s}$$

Moments and MGF

The nth moment of X equals the order-n derivative of $M_X(s)$ evaluated at $s = 0$.

$$\mathbf{E}[X^n] = \frac{d^n}{ds^n} M_X(s) \Big|_{s=0}$$

The order-n derivative of $M_X(s)$ is

$$\frac{d^n}{ds^n} M_X(s) = \frac{d^n}{ds^n} \int e^{sx} f_X(x) dx = \int x^n e^{sx} f_X(x) dx$$

At $s = 0$

$$\frac{d^n}{ds^n} M_X(s) \Big|_{s=0} = \int x^n f_X(x) dx = \mathbf{E}[X^n]$$

Example (4.27 Exponential)

For $X \sim \textbf{Exp}(\lambda)$, find the first/second moments of X via $M_X(s)$.

$$M_X(s) = \frac{\lambda}{\lambda - s}$$

$$\mathbf{E}[X] = \frac{d}{ds} M_X(s)\bigg|_{s=0} = \frac{\lambda}{(\lambda - s)^2}\bigg|_{s=0} = \frac{1}{\lambda}$$

$$\mathbf{E}\left[X^2\right] = \frac{d^2}{ds^2} M_X(s)\bigg|_{s=0} = \frac{2\lambda}{(\lambda - s)^3}\bigg|_{s=0} = \frac{2}{\lambda^2}$$

They are consistent with the results we found earlier.

Inversion property*

Let X be random variable. Suppose the MGF $M_X(s)$ is finite for all s in some interval $[-a, a]$. Then the distribution of X is uniquely decided by $M_X(s)$. When $M_X(s)$ can be recognized, the distribution of X can be named.

$$\begin{cases} p_X(x) \xrightarrow{\text{transform}} M_X(s) = \sum_{x \in \mathcal{X}} e^{sx} p_X(x) \\ M_X(s) \xrightarrow{\text{inverse transform}} p_X(x) \end{cases}$$

$$\begin{cases} f_X(x) \xrightarrow{\text{transform}} M_X(s) = \int e^{sx} f_X(x) dx \\ M_X(s) \xrightarrow{\text{inverse transform}} f_X(x) \end{cases}$$

Example (4.28)

Suppose X has MGF

$$M_X(s) = \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s}$$

Find the distribution of X.

$$\begin{aligned} M_X(s) &= \frac{1}{4}e^{-s} + \frac{1}{2} + \frac{1}{8}e^{4s} + \frac{1}{8}e^{5s} \\ &= \frac{1}{4}e^{s\cdot(-1)} + \frac{1}{2}e^{s\cdot 0} + \frac{1}{8}e^{s\cdot 4} + \frac{1}{8}e^{s\cdot 5} \\ &= \sum_{x\in\mathcal{X}} p_X(x)e^{sx} \end{aligned}$$

$$p_X(-1) = \frac{1}{4},\ p_X(0) = \frac{1}{2},\ p_X(4) = \frac{1}{8},\ p_X(5) = \frac{1}{8}$$

Example (4.29)

Suppose X has MGF

$$M_X(s) = \frac{pe^s}{1-(1-p)e^s}$$

Find the distribution of X.

$$\begin{aligned} M_X(s) &= \frac{pe^s}{1-(1-p)e^s} \\ &= pe^s\left(1+(1-p)e^s+(1-p)^2e^{2s}+\dots\right) \\ &= \sum_{x=1}^{\infty} p(1-p)^{x-1}e^{sx} = \sum_{x=1}^{\infty} p_X(x)e^{sx} \end{aligned}$$

$$\Rightarrow\ p_X(x) = p(1-p)^{x-1},\ x=1,2,\dots$$

Example (4.30)

A bank has 2 fast tellers and 1 slow teller. The time for a fast (resp. slow) teller to assist a customer is **Exp**(6) (resp. **Exp**(4)). Let X be the time a customer is assisted by a teller, each chosen with probability 1/3. Find the PDF and the MGF of X.

Let $A = \{$a fast teller is chosen$\}$.

$$f_X(x) = P(A)f_{X|A}(x) + P(A^c)f_{X|A^c}(x)$$

$$\begin{aligned} M_X(s) &= \int e^{sx} f_X(x)dx \\ &= \int e^{sx}\left(P(A)f_{X|A}(x) + P(A^c)f_{X|A^c}(x)\right)dx \\ &= P(A)M_{X|A}(s) + P(A^c)M_{X|A^c}(s) \\ &= \frac{2}{3}\frac{6}{6-s} + \frac{1}{3}\frac{4}{4-s} \end{aligned}$$

Sum MGF

The MGF of the sum of independent random variables is the product of the MGFs of the individual random variables.

Let X and Y be independent random variables and $Z = X + Y$. Then

$$\begin{aligned} M_Z(s) &= \mathbf{E}\left[e^{sZ}\right] = \mathbf{E}\left[e^{s(X+Y)}\right] = \mathbf{E}\left[e^{sX}e^{sY}\right] \\ &= \mathbf{E}\left[e^{sX}\right]\ \mathbf{E}\left[e^{sY}\right] \\ &= M_X(s)M_Y(s) \end{aligned}$$

Let $X_1, \ldots, X_n$ be independent and $Z = X_1 + \cdots + X_n$. Then

$$M_Z(s) = M_{X_1}(s) \ldots M_{X_n}(s)$$

Example (4.31 Bernoulli sum)

Find the MGF of $Z \sim \textbf{Bin}(n, p)$.

Z is the sum of n **iid** Bernoulli, i.e. $Z = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are independent $\textbf{Ber}(p)$. The MGF of X_i is

$$M_{X_i}(s) = (1-p)e^{s\cdot 0} + pe^{s\cdot 1}$$
$$= (1-p) + pe^s$$

So the MGF of Z is

$$M_Z(s) = M_{X_1}(s) \ldots M_{X_n}(s)$$
$$= (1 - p + pe^s)^n$$

Example (4.32 Poisson sum)

For independent $X \sim \textbf{Poi}(\lambda)$ and $Y \sim \textbf{Poi}(\mu)$, find the MGF of $Z = X + Y$.

The MGF of Z is

$$M_Z(s) = M_X(s)M_Y(s) = e^{\lambda(e^s-1)}e^{\mu(e^s-1)} = e^{(\lambda+\mu)(e^s-1)}$$

It shows $Z \sim \textbf{Poi}(\lambda + \mu)$. Alternatively, the PMF of Z can be derived by convolution

$$\begin{aligned} p_Z(z) &= \sum_{x=0}^{z} p_X(x)p_Y(z-x) = \sum_{x=0}^{z} \left(e^{-\lambda}\frac{\lambda^x}{x!}\right)\left(e^{-\mu}\frac{\mu^{z-x}}{(z-x)!}\right) \\ &= e^{-(\lambda+\mu)}\frac{1}{z!}\sum_{x=0}^{z}\frac{z!}{x!(z-x)!}\lambda^x\mu^{z-x} \\ &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^z}{z!} \end{aligned}$$

which also shows $Z \sim \textbf{Poi}(\lambda + \mu)$.

Example (4.33 Normal sum)

For independent $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$, find the MGF of $Z = X + Y$.

The MGF of Z is

$$\begin{aligned} M_Z(s) &= M_X(s) M_Y(s) \\ &= e^{\frac{1}{2}\sigma_x^2 s^2 + \mu_x s} \, e^{\frac{1}{2}\sigma_y^2 s^2 + \mu_y s} \\ &= e^{\frac{1}{2}(\sigma_x^2 + \sigma_y^2) s^2 + (\mu_x + \mu_y) s} \end{aligned}$$

This shows

$$(X + Y) \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Definition (random sum)

Let $X_1, X_2, \ldots$ be random variables and N be non-negative integer random variable (e.g. geometric, binomial, Poisson). Then

$$Y = X_1 + \cdots + X_N$$

is a random sum.

- A random sum consists of 2 levels of randomness
- Consider the example of Hi-Life. The total sale is a random sum

  $$Y = X_1 + \cdots + X_N$$

  where N is the number of customer and X_i is the purchase amount of customer i

Mean and variance

Let $X_1, X_2, \ldots$ be independent and identically distributed (abbr. **iid**) random variables, N be a non-negative integer random variable independent of X_i, and $Y = X_1 + \cdots + X_N$. Then

$$\mathbf{E}[Y] = \mathbf{E}[X_i]\mathbf{E}[N]$$

$$\textbf{var}(Y) = \textbf{var}(X_i)\,\mathbf{E}[N] + \mathbf{E}^2[X_i]\,\textbf{var}(N)$$

Conditional on N, the distribution of Y is clear. Thus we exploit N in iterated expectation and total variance.

$$\begin{aligned}
\mathbf{E}[Y] &= \mathbf{E}[\mathbf{E}[Y|N]] = \mathbf{E}[\mathbf{E}[X_1 + \cdots + X_N|N]] \\
&= \mathbf{E}[\mathbf{E}[X_1|N] + \cdots + \mathbf{E}[X_N|N]] \\
&= \mathbf{E}[N\,\mathbf{E}[X_i]] \\
&= \mathbf{E}[X_i]\mathbf{E}[N] \\
\textbf{var}(Y) &= \mathbf{E}[\textbf{var}(Y|N)] + \textbf{var}(\mathbf{E}[Y|N]) \\
&= \mathbf{E}[N\,\textbf{var}(X_i|N)] + \textbf{var}(N\,\mathbf{E}[X_i|N]) \\
&= \textbf{var}(X_i)\,\mathbf{E}[N] + \mathbf{E}^2[X_i]\,\textbf{var}(N)
\end{aligned}$$

MGF

Let $X_1, X_2, \ldots$ be **iid** random variables, N be a non-negative integer random variable independent of X_i, and $Y = X_1 + \cdots + X_N$.

$$M_Y(s) = M_N(\log M_{X_i}(s))$$

$$\begin{aligned} M_Y(s) &= \mathbf{E}\left[e^{sY}\right] = \mathbf{E}\left[\mathbf{E}\left[e^{sY} \middle| N\right]\right] \\ &= \mathbf{E}\left[(M_{X_i}(s))^N\right] \\ &= \sum_n (M_{X_i}(s))^n p_N(n) \end{aligned}$$

$$M_N(s) = \mathbf{E}\left[e^{sN}\right] = \sum_n e^{sn} p_N(n)$$

$$\Rightarrow \quad M_N(\log M_{X_i}(s)) = \sum_n (M_{X_i}(s))^n \, p_N(n) = M_Y(s)$$

Example (4.34 Uniform random sum)

A remote village has 3 gas stations. Each gas station is open on any given day with probability $\frac{1}{2}$, independent of the others. The amount of gas available in an open station is **Uni**$(0, 1000)$ gallons. Let Y be the total amount of gas available at the open gas stations in one day. Find the MGF of Y.

Let N be the number of open stations and X_i be the amount of gas available at the ith open station. Then the total amount of gas is a random sum. The MGF of Y depends on the MGF of X_i and N. The MGF of $N \sim$ **Bin**$(3, \frac{1}{2})$ is

$$M_N(s) = \left(\frac{1}{2} + \frac{1}{2}e^s\right)^3 = \frac{1}{8}\left(1 + e^s\right)^3$$

The MGF of $X_i \sim \textbf{Uni}(0, 1000)$ is

$$M_{X_i}(s) = \mathbf{E}\left[e^{sX_i}\right] = \int e^{sx} f_{X_i}(x)dx = \int_0^{1000} e^{sx} \frac{1}{1000} dx$$
$$= \frac{e^{1000s} - 1}{1000s}$$

The MGF of $Y = X_1 + \cdots + X_N$ is

$$\begin{aligned} M_Y(s) &= M_N(\log M_{X_i}(s)) \\ &= \frac{1}{8}\left(1 + e^{\log M_{X_i}(s)}\right)^3 \\ &= \frac{1}{8}(1 + M_{X_i}(s))^3 \\ &= \frac{1}{8}\left(1 + \frac{e^{1000s} - 1}{1000s}\right)^3 \end{aligned}$$

Example (4.35 Exponential random sum)

Jane is looking for **Great Expectations**. In a bookstore, Jane spends a random amount of time, exponentially distributed with parameter λ, and she either finds a copy with probability p or try another bookstore. Find the mean, the variance, and the PDF of the total time Jane will spend until she finds the book.

Let N be the number of bookstores Jane visits and X_i is the time she spends at the ith bookstore. Then the total time spent by Jane is a random sum

$$Y = X_1 + \cdots + X_N$$

The PDF of Y can be found via $M_Y(s)$, which in turn can be found via $M_{X_i}(s)$ and $M_N(s)$.

$X_i \sim$ **Exp**(λ) and $N \sim$ **Geo**(p).

$$M_{X_i}(s) = \frac{\lambda}{\lambda - s}, \; M_N(s) = \frac{pe^s}{1-(1-p)e^s}$$

$$\Rightarrow \; M_Y(s) = M_N(\log M_{X_i}(s)) = \frac{pM_{X_i}(s)}{1-(1-p)M_{X_i}(s)}$$

$$= \frac{p \, \frac{\lambda}{\lambda - s}}{1-(1-p) \, \frac{\lambda}{\lambda-s}} = \frac{p\lambda}{p\lambda - s}$$

This shows $Y \sim$ **Exp**$(p\lambda)$. It follows that

$$f_Y(y) = (p\lambda)e^{-(p\lambda)y}u(y), \quad \mathbf{E}[Y] = \frac{1}{p\lambda}, \quad \textsf{\textbf{var}}(Y) = \frac{1}{p^2\lambda^2}$$

Example (4.36 Geometric random sum)

Suppose $X_i \sim \textbf{Geo}(q)$, $N \sim \textbf{Geo}(p)$ and $Y = X_1 + \cdots + X_N$. Then

$$Y \sim \textbf{Geo}(pq)$$

$$M_{X_i}(s) = \frac{qe^s}{1-(1-q)e^s}, \;\; M_N(s) = \frac{pe^s}{1-(1-p)e^s}$$

$$\Rightarrow \;\; M_Y(s) = M_N(\log M_{X_i}(s)) = \frac{pM_{X_i}(s)}{1-(1-p)M_{X_i}(s)}$$

$$= \frac{p\,\frac{qe^s}{1-(1-q)e^s}}{1-(1-p)\frac{qe^s}{1-(1-q)e^s}} = \frac{pq\,e^s}{1-(1-pq)e^s}$$

So $Y \sim \textbf{Geo}(pq)$.

Example 4.35 *

Example 4.35 can be solved by total probability. For $y > 0$

$$
\begin{aligned}
f_Y(y) &= \sum_{n=1}^{\infty} p_N(n) f_{Y|N=n}(y) \\
&= \sum_{n=1}^{\infty} p(1-p)^{n-1} \overbrace{\frac{\lambda^n y^{n-1} e^{-\lambda y}}{(n-1)!}}^{\text{PDF of the } n\text{th arrival time}} \\
&= p\lambda e^{-\lambda y} \sum_{n=1}^{\infty} (1-p)^{n-1} \frac{\lambda^{n-1} y^{n-1}}{(n-1)!} \\
&= p\lambda e^{-\lambda y} e^{(1-p)\lambda y} \\
&= p\lambda e^{-p\lambda y}
\end{aligned}
$$

Example 4.36 *

Example 4.36 can also be solved by total probability. For $y \in \mathbb{N}$

$$
\begin{aligned}
p_Y(y) &= \sum_{n=1}^{\infty} p_N(n) p_{Y|N=n}(y)\, \mathbb{I}_{y \geq n} \\
&= \sum_{n=1}^{y} p(1-p)^{n-1} \overbrace{\binom{y-1}{n-1} q^n (1-q)^{y-n}}^{\text{PMF of the } n\text{th arrival time}} \\
&= pq \sum_{n'=0}^{y-1} \binom{y-1}{n'} (1-p)^{n'} q^{n'} (1-q)^{y-1-n'} \\
&= pq((1-p)q + (1-q))^{y-1} \\
&= pq(1-pq)^{y-1}
\end{aligned}
$$

Summary 1

2-step method for the distribution of $Y = g(X)$

$$P(Y \leq y) = P(g(X) \leq y) = \sum_i P(l_i(y) \leq X \leq u_i(y))$$

$$F_Y(y) = \sum_i F_X(u_i(y)) - F_X(l_i(y))$$

Covariance and correlation

$$\textbf{cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

$$\textbf{Corr}(X, Y) = \frac{\textbf{cov}(X, Y)}{\sqrt{\textbf{var}(X)}\sqrt{\textbf{var}(Y)}}$$

Summary 2

Conditional expectation and variance

$$\mathbf{E}[X|Y] = g(Y)$$
$$\mathbf{var}(X|Y) = h(Y)$$

Law of iterated expectation

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$$

Law of total variance

$$\mathbf{var}(X) = \mathbf{var}(\mathbf{E}[X|Y]) + \mathbf{E}[\mathbf{var}(X|Y)]$$

Summary 3

MGF

$$M_X(s) = \mathbf{E}\left[e^{sX}\right]$$

$$\mathbf{E}[X^n] = \frac{d^n}{ds^n} M_X(s)\Big|_{s=0}$$

Random sum of **iid** random variables

$$Y = X_1 + \cdots + X_N$$

$$\mathbf{E}[Y] = \mathbf{E}[N]\mathbf{E}[X_i]$$

$$\textbf{var}(Y) = \textbf{var}(X_i)\,\mathbf{E}[N] + \mathbf{E}^2[X_i]\,\textbf{var}(N)$$

$$M_Y(s) = M_N(\log M_{X_i}(s))$$