San José State University

Math 263: Stochastic Processes

Probability review

Dr. Guangliang Chen

This lecture is based on the following textbook sections:

- Sections 3.2 - 3.5
- Section 5.2

Outline of the presentation

- Poisson, Exponential and Gamma distributions
- Conditional distribution and expectation

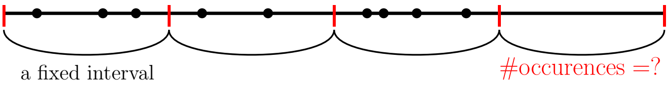
Hw1: Assigned in Canvas

The Poisson distribution

Recall that the Poisson distribution has the following pmf

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, \ldots$$

It can be used to model the number of occurrences of a rare event over a time/space interval of fixed length with rate λ.



Theorem 0.1. If $X \sim \text{Pois}(\lambda)$, then $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$.

The Exponential distribution

Recall also that a random variable X is said to have an exponential distribution with parameter λ if it has the following pdf:

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

It is useful for modeling waiting time for one occurrence of a rare event.

Theorem 0.2. If $X \sim \mathrm{Exp}(\lambda)$, then

$$F(x) = 1 - e^{-\lambda x}, \qquad \bar{F}(x) = e^{-\lambda x}, \qquad x > 0$$

$$\mathrm{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

An important fact about this distribution is the **memoryless property**:

$$P(X > s + t \mid X > s) = P(X > t), \quad \forall\, s, t > 0$$

To see this (again),

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t, \cancel{X > s})}{P(X > s)} \\ &= \frac{\bar{F}(s+t)}{\bar{F}(s)} = \frac{e^{-(s+t)}}{e^{-s}} \\ &= e^{-t} = \bar{F}(t) \\ &= P(X > t). \end{aligned}$$

Theorem 0.3. If $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$ are independent, then

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Proof.

$$\begin{aligned} P(X_1 < X_2) &= \int_0^\infty \int_0^{x_2} \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &= \int_0^\infty (1 - e^{-\lambda_1 x_2}) \lambda_2 e^{-\lambda_2 x_2} \, \mathrm{d}x_2 \\ &= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2} \, \mathrm{d}x_2 - \int_0^\infty \lambda_2 e^{-(\lambda_1 + \lambda_2) x_2} \, \mathrm{d}x_2 \\ &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \end{aligned}$$

Theorem 0.4. If $X_i \sim \text{Exp}(\lambda_i), i = 1, \ldots, n$ are independent, then

$$\min_{1 \le i \le n} X_i \sim \text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Proof. For any $x > 0$,

$$\begin{aligned} P\left(\min_{1 \le i \le n} X_i > x\right) &= P(X_1 > x, \ldots, X_n > x) \\ &= \prod_{i=1}^{n} P(X_i > x) \\ &= \prod_{i=1}^{n} e^{-\lambda_i x} = e^{-(\sum \lambda_i) x}. \end{aligned}$$

Consider a positive (continuous) random variable X with pdf $f(x)$ and cdf $F(x)$. The hazard rate function of X is defined as follows:

Def 0.1 (Hazard rate function)**.**

$$r(t) = \frac{f(t)}{1 - F(t)}, \quad t > 0$$

To understand the meaning of $r(t)$, suppose X represents the operation time of a machine (in hours). The probability that the machine will break down during a tiny time period right after it has lasted for t hours is

$$P(X \in (t, t + \Delta t) \mid X > t) = \frac{P(t < X < t + \Delta t, \cancel{X > t})}{P(X > t)} \approx \frac{f(t)\Delta t}{1 - F(t)} = r(t)\Delta t.$$

If $X \sim \text{Exp}(\lambda)$, then

$$r(t) = \frac{\lambda e^{-\lambda t}}{1-(1-e^{-\lambda t})} = \lambda \quad \text{(constant failure rate)}.$$

Given a hazard rate function $r(t)$, we may uniquely reconstruct the cdf of the random variable X. First, rewrite

$$r(t) = \frac{\frac{d}{dt}F(t)}{1-F(t)} = \frac{d}{dt}[-\log(1-F(t))]$$

Integrating both sides from 0 to t gives that

$$\int_0^t r(s)\,ds = -\log(1-F(t))$$

From this we obtain that

$$F(t) = 1 - e^{-\int_0^t r(s)\,ds}.$$

The Gamma distribution

Def 0.2. A random variable X is said to have a Gamma distribution, with parameters α and λ, if its pdf has the following form

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0$$

where

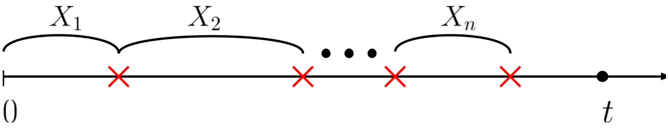
$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \,\mathrm{d}x, \quad \alpha > 0$$

Remark. If $\alpha = n$ is an integer, then $\Gamma(n) = (n-1)!$. So it is a generalization of the factorial function from positive integers to positive real numbers.

The Poisson-Exponential-Gamma scheme

Suppose a rare event occurs with rate λ over time.

- For any $t > 0$, let $N(t)$ be the total number of occurrences of this event by time t. Then $N(t) \sim \text{Pois}(\lambda t)$.
- For any positive integer n, let $X_i, 1 \le i \le n$ represent the waiting time for the ith occurrence of the event (after the last occurrence). Then $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(\lambda)$.



Let T be the total amount of waiting time for n occurrences of the event:

$$T = \sum_{i=1}^{n} X_i.$$

Then

$$T \sim \text{Gamma}(n, \lambda).$$

Proof. For any fixed $t > 0$, the cdf of T is

$$F_T(t) = P(T < t) = P(N(t) \geq n) = \sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Differentiating $F_T(t)$ with respect to t gives that

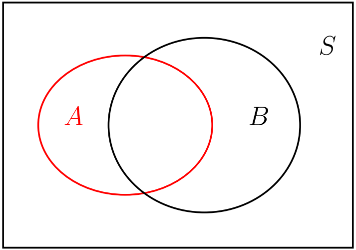
$$f_T(t) = \frac{\lambda(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!}$$

This shows that $T \sim \mathrm{Gamma}(n, \lambda)$. $\square$

Conditional distributions

For two events $A, B \subset S$ with $P(A) > 0$, the conditional probability of B given A is defined as

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$$



For two random variables X, Y that have a joint distribution, their conditional distributions can be defined similarly.

Two discrete random variables

The conditional pmf of X given $Y = y$ is defined as

$$p_{X|Y}(x \mid \underbrace{y}_{\text{fixed}}) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{P(X = x, Y = y)}{P(Y = y)}$$

from which one can compute the conditional cdf, expectation and variance.

Example 0.1. Let $X \sim \text{Pois}(\lambda_1), Y \sim \text{Pois}(\lambda_2)$ be independent random variables. Find $\text{E}(X \mid X + Y = n)$.

Solution. We start by computing the conditional pmf of X given $X+Y=n$: For each $x=0,1,\ldots,n$,

$$P(X=x \mid X+Y=n) = \frac{P(X=x, X+Y=n)}{P(X+Y=n)} = \frac{P(X=x, Y=n-x)}{P(X+Y=n)}$$
$$= \frac{P(X=x)P(Y=n-x)}{P(X+Y=n)} = \binom{n}{x}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-x}$$

This shows that

$$X \mid X+Y=n \quad \sim \quad B\left(n,\ p=\frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

Therefore,

$$\mathrm{E}[X \mid X+Y=n] = \frac{n\lambda_1}{\lambda_1+\lambda_2}.$$

Two continuous random variables

The conditional pdf of X given $Y = y$ is defined as

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

from which one can compute the conditional cdf, expectation and variance.

Example 0.2. Consider the following joint distribution

$$f(x,y) = \frac{1}{2}xy, \quad 0 < x < y < 2$$

Find $\mathrm{E}(X \mid Y = y)$.

Solution. By direct calculation,

$$
\begin{aligned}
f_Y(y) &= \int_0^y \frac{1}{2} x y \, \mathrm{d}x = \frac{1}{4} y^3, \quad 0 < y < 2 \\
f_{X|Y}(x \mid y) &= \frac{\frac{1}{2} x y}{\frac{1}{4} y^3} = \frac{2}{y^2} x, \quad 0 < x < y \\
\mathrm{E}(X \mid Y = y) &= \int_0^y x \cdot \frac{2}{y^2} x \, \mathrm{d}x = \frac{2}{3} y. \qquad \square
\end{aligned}
$$

Let $\mathrm{E}(X \mid Y)$ be the expression of $\mathrm{E}(X \mid Y = y)$ with each y replaced by Y. In the above example,

$$
\mathrm{E}(X \mid Y) = \frac{2}{3} Y.
$$

Note that $\mathrm{E}(X \mid Y)$ is a random variable (dependent on Y).

Theorem 0.5. For any two random variables X, Y with a joint distribution,

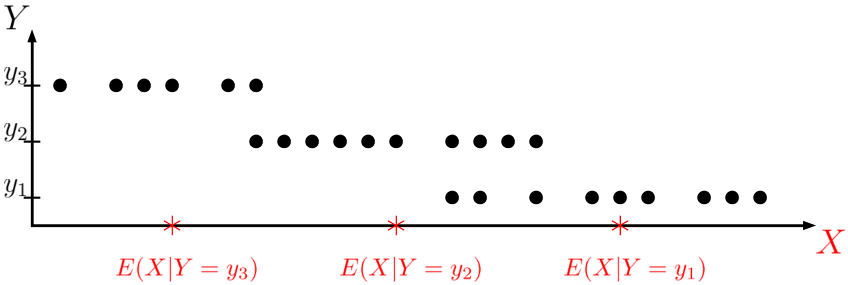
$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid Y)).$$

Proof. We prove this result for the case of two discrete random variables:

$$\begin{aligned}
\mathrm{E}(\mathrm{E}(X \mid Y)) &= \sum_y \mathrm{E}(X \mid Y = y) P(Y = y) \\
&= \sum_y \sum_x x\, P(X = x \mid Y = y) P(Y = y) \\
&= \sum_x x \sum_y P(X = x \mid Y = y) P(Y = y) \\
&= \sum_x x\, P(X = x) = \mathrm{E}(X).
\end{aligned}$$

Remark. When Y is discrete, this formula can be interpreted as follows:

$$\underbrace{\mathrm{E}(X)}_{\text{overall mean}} = \underbrace{\mathrm{E}(\ \underbrace{\mathrm{E}(X \mid Y)}_{\text{group mean}}\ )}_{\text{further averaged across groups}}.$$



Example 0.3 (cont'd)**.**

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid Y)) = \mathrm{E}\left(\frac{2}{3}Y\right) = \frac{2}{3}\int_0^2 y \cdot \frac{1}{4}y^3 \,\mathrm{d}y = \frac{2}{3}\cdot\frac{8}{5} = \frac{16}{15}.$$

Verify this result by using the marginal pdf of X instead.

Example 0.4. Let $X_1, X_2, \ldots$ be a sequence of iid random variables with the same mean $\mu = \mathrm{E}(X_i)$ and variance $\sigma^2 = \mathrm{Var}(X_i)$. Let N be a positive, integer-valued random variable that is independent of all X_i. Define

$$S = \sum_{i=1}^{N} X_i$$

which is a compound random variable. Prove that

$$\mathrm{E}(S) = \mu \cdot \mathrm{E}(N).$$

Proof.

$$\mathrm{E}(S) = \mathrm{E}(\mathrm{E}(S \mid N)) = \mathrm{E}(\mu N) = \mu \cdot \mathrm{E}(N).$$

Example 0.5. Let

$$X_i \sim \mathrm{Exp}(\lambda_i), \quad 1 \le i \le n$$

be independent random variables, and N another random variable that is independent of all X_i and has the following distribution

$$P(N = i) = p_i, \quad 1 \le i \le n.$$

Let

$$Y = X_N,$$

which is called a hyperexponential random variable. Find $\mathrm{E}(Y)$.

Solution.

$$\mathrm{E}(Y) = \mathrm{E}(\mathrm{E}(X_N \mid N)) = \mathrm{E}\left(\frac{1}{\lambda_N}\right) = \sum_{i=1}^{n} \frac{1}{\lambda_i} \cdot p_i = \sum_{i=1}^{n} \frac{p_i}{\lambda_i}.$$

We can also determine the density of Y: For any $t > 0$,

$$\begin{aligned} P(Y > t) = P(X_N > t) &= \sum_{i=1}^{n} P(X_N > t \mid N = i) P(N = i) \\ &= \sum_{i=1}^{n} P(X_i > t \mid \cancel{N = i}) P(N = i) \\ &= \sum_{i=1}^{n} e^{-\lambda_i t} p_i. \end{aligned}$$

It follows that

$$F_Y(t) = 1 - \sum_{i=1}^{n} p_i e^{-\lambda_i t}$$

and

$$f_Y(t) = \sum_{i=1}^{n} p_i \lambda_i e^{-\lambda_i t}.$$

The hazard rate function is

$$r(t) = \frac{f(t)}{1 - F(t)} = \frac{\sum p_i \lambda_i e^{-\lambda_i t}}{\sum p_i e^{-\lambda_i t}} \xrightarrow{t\to\infty} \min_{1\le i\le n} \lambda_i.$$

Theorem 0.6. For any random variables X, Y that have a joint distribution,

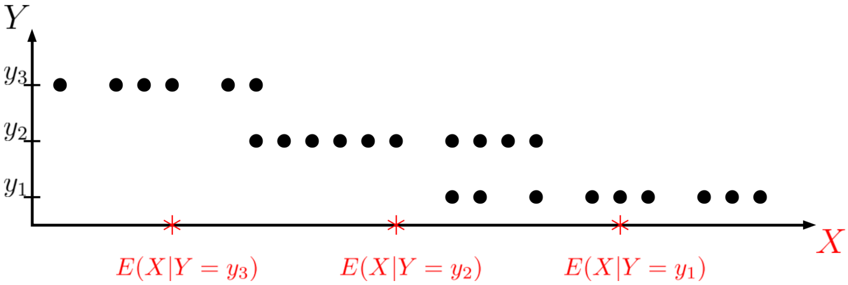
$$\mathrm{Var}(X) = \mathrm{E}(\mathrm{Var}(X \mid Y)) + \mathrm{Var}(\mathrm{E}(X \mid Y)).$$

Proof.

$$\mathrm{E}(\mathrm{Var}(X \mid Y)) = \mathrm{E}\left(\mathrm{E}(X^2 \mid Y) - \mathrm{E}(X \mid Y)^2\right) = \mathrm{E}(X^2) - \mathrm{E}\left(\mathrm{E}(X \mid Y)^2\right)$$

$$\mathrm{Var}(\mathrm{E}(X \mid Y)) = \mathrm{E}\left(\mathrm{E}(X \mid Y)^2\right) - (\mathrm{E}(\mathrm{E}(X \mid Y)))^2 = \mathrm{E}\left(\mathrm{E}(X \mid Y)^2\right) - (\mathrm{E}(X))^2.$$

Remark. When Y is discrete, this formula can be interpreted as a decomposition of the total variance of X into within-group and between-group variances:



Example 0.6 (Compound random variable, cont'd)**.** For

$$S = \sum_{i=1}^{N} X_i,$$

we have

$$\begin{aligned} E(S \mid N) &= \mu N & &\longrightarrow \mathrm{Var}(E(S \mid N)) = \mu^2 \mathrm{Var}(N) \\ \mathrm{Var}(S \mid N) &= \mathrm{Var}\left[\sum_{i=1}^{N} X_i \mid N\right] = \sigma^2 N & &\longrightarrow E(\mathrm{Var}(S \mid N)) = \sigma^2 E(N) \end{aligned}$$

Accordingly,

$$\mathrm{Var}(S) = \mu^2 \mathrm{Var}(N) + \sigma^2 E(N)$$