Chapter 3 Summary

Expectation of Discrete Random Variables

- If X is a discrete random variable, then

$$\mathsf{E}(X) = \mu_X = \sum_{x \in \mathbb{R}} x\mathsf{P}(X = x)$$

  is the *expected value* of the random variable X.

- If X has distinct possible values x_i with corresponding probabilities p_i, then

$$\mathsf{E}(X) = \sum_i x_i p_i$$

- The expected value is the *weighted average* of the possible values of X where the weights are the probabilities.

- The expected value measures the center of a distribution by specifying the *balancing point*.

- Expectation is a *linear* operator.

$$\mathsf{E}(aX) = a\mathsf{E}(X), \quad \mathsf{E}(X+Y) = \mathsf{E}(X) + \mathsf{E}(Y)$$

  which is true for *any* random variables X and Y.

- Expectations of *functions of random variables* are the weighted average of the functions of the possible values of X weighted by their probabilities.

$$\mathsf{E}(g(X)) = \sum_{x \in \mathbb{R}} g(X)\mathsf{P}(X = x)$$

- Infinite expectation. Expectations can be *infinite* or *undefined*.
  - If $\sum_{x \in Real, x>0} x\mathsf{P}(X = x) = +\infty$ and $-\sum_{x \in Real, x<0} x\mathsf{P}(X = x) < +\infty$, then $\mathsf{E}(X) = +\infty$.
  - If $\sum_{x \in Real, x>0} x\mathsf{P}(X = x) < +\infty$ and $-\sum_{x \in Real, x<0} x\mathsf{P}(X = x) = +\infty$, then $\mathsf{E}(X) = -\infty$.
  - If $\sum_{x \in Real, x>0} x\mathsf{P}(X = x) = +\infty$ and $-\sum_{x \in Real, x<0} x\mathsf{P}(X = x) = +\infty$, then $\mathsf{E}(X)$ is not defined.

- If X and Y are *independent*, then

$$\mathsf{E}(g(X)h(Y)) = \mathsf{E}(g(X))\mathsf{E}(h(Y))$$

  for functions g and h. In particular,

$$\mathsf{E}(XY) = \mathsf{E}(X)\mathsf{E}(Y) .$$

- (*Monotonicity of Expectation*): If $X \geq Y$, then $\mathsf{E}(X) \geq \mathsf{E}(Y)$.

Expectation of Absolutely Continuous Random Variables

- If X is an *absolutely continuous random variable* with density f_X, then

$$\mathsf{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x .$$

- Expectation of absolutely continuous random variables follow the same proporties as discrete random variables:
  - *linearity*, $\mathsf{E}(aX + bY) = a\mathsf{E}(X) + b\mathsf{E}(Y)$;
  - *functions of random variables*, $\mathsf{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$;
  - if *independent*, then $\mathsf{E}(g(X)h(Y)) = \mathsf{E}(g(X))\mathsf{E}(h(Y))$;
  - expectations can be *infinite* or *undefined*;
  - *monotonicity*, if $X \geq Y$, then $\mathsf{E}(X) \geq \mathsf{E}(Y)$.

Variance

- The *variance* of a random variable measures the *spread* of the probability distribution and is the *expected squared distiance of the random variable from its mean*, $\mathsf{E}(X) = \mu_X$.

$$\mathsf{Var}(X) = \sigma_X^2 = \mathsf{E}((X - \mu_X)^2) .$$

- The variance is only defined when the mean is defined, and the variance can be infinite.

- The variance has these properties:
  - $\mathsf{Var}(X) \geq 0$;
  - $\mathsf{Var}(aX + b) = a^2\mathsf{Var}(X)$ for $a, b \in \mathbb{R}$;
  - $\mathsf{Var}(X) = \mathsf{E}(X^2) - \mu_X^2$;
  - $\mathsf{Var}(X) \leq \mathsf{E}(X^2)$.

Means and Variances of Important Distributions

- Discrete distributions.
  - **Degenerate distributions:** $\mathsf{P}(X = c) = 1$ for some c.
    $\mathsf{E}(X) = c$, $\mathsf{Var}(X) = 0$.
  - **Bernoulli distributions:** $X \sim \text{Bernoulli}(\theta)$, $\mathsf{P}(X = 1) = \theta$, $\mathsf{P}(X = 0) = 1 - \theta$ where $0 < \theta < 1$ is a parameter.
    $\mathsf{E}(X) = \theta$, $\mathsf{Var}(X) = \theta(1 - \theta)$.
  - **Binomial distributions:** $X \sim \text{Binomial}(n, \theta)$.

$$\mathsf{P}(X = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \quad \text{for } k = 0, 1, \ldots, n .$$

    $\mathsf{E}(X) = n\theta$, $\mathsf{Var}(X) = n\theta(1 - \theta)$.
  - **Geometric distributions:** $X \sim \text{Geometric}(\theta)$.

$$\mathsf{P}(X = k) = \theta(1 - \theta)^k \quad \text{for } k = 0, 1, 2, \ldots .$$

    $\mathsf{E}(X) = (1 - \theta)/\theta$, $\mathsf{Var}(X) = (1 - \theta)/\theta^2$.

- **Negative Binomial distributions:** $X \sim$ Negative Binomial(r, θ).

$$\mathsf{P}(X = k) = \binom{k+r-1}{r-1} \theta^r (1-\theta)^k \quad \text{for } k = 0, 1, 2, \ldots .$$

$\mathsf{E}(X) = r(1-\theta)/\theta$, $\mathsf{Var}(X) = r(1-\theta)/\theta^2$.

- **Poisson distributions:** $X \sim$ Poisson(λ). For parameter $\lambda > 0$,

$$\mathsf{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots .$$

$\mathsf{E}(X) = \lambda$, $\mathsf{Var}(X) = \lambda$.

- **Hypergeometric distributions:** $X \sim$ Hypergeometric(N, M, n).

$$\mathsf{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},$$

where $k = \max\{0, n - N + M\}, \ldots, \min\{M, n\}$.
$\mathsf{E}(X) = nM/N$, $\mathsf{Var}(X) = n(M/N)((N - M)/N)((N-n)/(N-1))$.

- Absolutely Continuous Distributions.
  - **Uniform distributions:** $X \sim$ Uniform(L, R).

$$f(x) = 1/(R - L) \quad \text{for } L < x < R .$$

$\mathsf{E}(X) = (L+R)/2$, $\mathsf{Var}(X) = (R-L)^2/12$.

  - **Exponential distributions:** $X \sim$ Exponential(λ). For parameter $\lambda > 0$, the density is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 .$$

$\mathsf{E}(X) = 1/\lambda$, $\mathsf{Var}(X) = 1/\lambda^2$.

  - **Gamma distributions:** $X \sim$ Gamma(α, λ), for shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$. The density is

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{for } x \geq 0,$$

where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$

is the *gamma function*, defined when $\alpha > 0$.
$\mathsf{E}(X) = \alpha/\lambda$, $\mathsf{Var}(X) = \alpha/\lambda^2$.

  - **Normal distributions:** $X \sim N(\mu, \sigma^2)$. The density is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-((x-\mu)/\sigma)^2/2} \quad (-\infty < x < \infty) .$$

$\mathsf{E}(X) = \mu$, $\mathsf{Var}(X) = \sigma^2$.

  - **Beta distributions:** $X \sim$ Beta(a, b). With parameters $a > 0, b > 0$, the density is

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad \text{for } 0 < x < 1 .$$

where the *beta function* $B(a, b)$ is defined for $a > 0, b > 0$ as

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} .$$

$\mathsf{E}(X) = a/(a+b)$, $\mathsf{Var}(X) = ab/((a+b)^2(a+b+1))$.

Covariance

- The *covariance* of random variables X and Y is

$$\mathsf{Cov}(X, Y) = \mathsf{E}((X - \mu_X)(Y - \mu_Y))$$

where $\mu_X = \mathsf{E}(X)$ and $\mu_Y = \mathsf{E}(Y)$.

- $$\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$$

- (*Linearity of Covariance*) For random variables X_i and Y_j and real numbers a_i and b_j,

$$\mathsf{Cov}\left(\sum_i a_i X_i, \sum_j b_j Y_j\right) = \sum_i \sum_j a_i b_j \mathsf{Cov}(X_i, Y_j) .$$

- (*Variance of Sums*)

$$\mathsf{Var}(\sum_i X_i) = \sum_i \mathsf{Var}(X_i) + 2\sum_{i<j} \mathsf{Cov}(X_i, X_j)$$

- The *correlation* between random variables X and Y is

$$\mathsf{Corr}(X, Y) = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}} = \frac{\mathsf{Cov}(X,Y)}{\mathsf{SD}(X)\mathsf{SD}(Y)}$$

provided that $0 < \mathsf{Var}(X), \mathsf{Var}(Y) < \infty$.

- The correlation has possible values $-1 \leq \mathsf{Corr}(X, Y) \leq 1$.

Probability Generating Functions

- If X is a random variable (usually discrete on 0, 1, 2, ...or a subset thereof), then $r(t) = \mathsf{E}(t^X)$ is the *probability generating function* of X.

- If X is a discrete random variable whose possible values are all nonnegative integers and if $r(t_0) < \infty$ for some $t_0 > 0$, then the kth derivative evaluated at 0 is related to the probability that $X = k$.

$$\mathsf{P}(X = k) = r^{(k)}(0)/k!$$

- (*Uniqueness*) If two discrete random variables whose possible values are all nonnegative integers have the same probability generating function which is finite for some $t_o > 0$, then the random variables have the same distribution.

- If X and Y are *independent* and $Z = X + Y$, then

$$r_Z(t) = r_X(t) r_Y(t) .$$

Moment Generating Functions

- If X is a random variable, then $m(t) = \mathsf{E}(e^{tX})$ is the *moment generating function* of X.
- The kth *moment* of X is the expected value of the kth power of X, $\mathsf{E}(X^k)$.
- If X is a random variable and if $m(t) < \infty$ for all $t \in (-t_0, t_0)$ for some $t_0 > 0$, then the kth derivative evaluated at 0 is equal to the kth moment of X.

$$m^{(k)}(0) = \mathsf{E}(X^k)$$

- (*Uniqueness*) If two random variables have the same moment generating function which is finite for all $t \in (-t_0, t_0)$ for some $t_0 > 0$, then the random variables have the same distribution.
- If X and Y are *independent* and $Z = X + Y$, then

$$m_Z(t) = m_X(t)m_Y(t) \ .$$

Characteristic Functions

- If X is a random variable, then $c(t) = \mathsf{E}(e^{itX})$ is the *characteristic function* of X where $i = \sqrt{-1}$.
- The characteristic function is defined for all distributions, namely $c_X(t) < \infty$ for all t for any random variable X.
- If X is a random variable, then the kth derivative of $c(t)$ evaluated at 0 is related to the kth moment of X.

$$c^{(k)}(0) = i^k \mathsf{E}(X^k)$$

- (*Uniqueness*) If two random variables have the same characteristic function then the random variables have the same distribution.
- If X and Y are *independent* and $Z = X + Y$, then

$$c_Z(t) = c_X(t)c_Y(t) \ .$$

Conditional Expectation

- *Conditional expectations* are expected values under conditional distributions, and follow all of the properties of expectations for unconditional distributions.
- For a random variable X and event A,

$$\begin{aligned} \mathsf{E}(X \mid A) &= \sum_{x \in \mathbb{R}} x \mathsf{P}(X = x \mid A) \quad \text{if } X \text{ is discrete} \\ &= \int_{-\infty}^{\infty} x f_{X|A}(x) \quad \text{if } X \text{ is absolutely continuous} \end{aligned}$$

For random variables X and Y,

$$\begin{aligned} \mathsf{E}(X \mid Y = y) &= \sum_{x \in \mathbb{R}} x \mathsf{P}(X = x \mid Y = y) \quad \text{if } X \text{ is discrete} \\ &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) \quad \text{if } X \text{ is absolutely continuous} \end{aligned}$$

- The conditional expectation $\mathsf{E}(X \mid Y)$ is a *function of* Y, and is thus a *random variable*.
- (*The Law of Total Expectation*) For any two random variables X and Y,

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

- The variance of a conditional distribution is also a random variable.

$$\mathsf{Var}(X \mid Y) = \mathsf{E}((X - \mathsf{E}(X \mid Y))^2 \mid Y)$$

- There is an expression for variance similar to the law of total expectation.

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

Inqualities

- If X is a nonnegative random variable with a finite mean and $a > 0$ is a real number, the *Markov inequality* puts a limit on the tail probability of the distribution.

$$\mathsf{P}(X \geq a) \leq \frac{\mathsf{E}(X)}{a}$$

- If X is a random variable with a finite variance and $a > 0$ is a real number, the *Chebychev inequality* puts a limit on the two-sided tail probability of the distribution.

$$\mathsf{P}(|X - \mu_X| \geq a) \leq \frac{\mathsf{Var}(X)}{a^2}$$

- The *Cauchy-Schwartz Inquality* limits the covariance of two random variables and implies that the correlation is between -1 and 1.

$$\mathsf{Cov}(X, Y) \leq \sqrt{\mathsf{Var}(X)\mathsf{Var}(Y)}$$

- A function g is a *convex function* if for $a < b$, between a and b the function g is less than or equal to a line that connects $g(a)$ and $g(b)$. Specifically, for $a < b$, if $x = (1-\lambda)a + \lambda b$ for $0 \leq \lambda \leq 1$, then $g(x) \leq (1-\lambda)g(a) + \lambda g(b)$.
- *Jensen's Inequality* involves expected values of convex functions of random variables. If g is a *convex function*, then

$$g(\mathsf{E}(X)) \leq \mathsf{E}(g(X)) \ .$$