Chapter 3 Summary

Expectation of Discrete Random Variables

- If $X$ is a discrete random variable, then
  \[ E(X) = \mu_X = \sum_{x \in \mathbb{R}} x P(X = x) \]
is the expected value of the random variable $X$.

- If $X$ has distinct possible values $x_i$ with corresponding probabilities $p_i$, then
  \[ E(X) = \sum_{i} x_i p_i \]

- 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.
  \[ E(aX) = aE(X), \quad E(X + Y) = E(X) + E(Y) \]
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.
  \[ E(g(X)) = \sum_{x \in \mathbb{R}} g(x) P(X = x) \]

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

- If $X$ and $Y$ are independent, then
  \[ E(g(X)h(Y)) = E(g(X))E(h(Y)) \]
for functions $g$ and $h$. In particular,
\[ E(XY) = E(X)E(Y) \].

(Monotonicity of Expectation): If $X \geq Y$, then $E(X) \geq E(Y)$.

Expectation of Absolutely Continuous Random Variables

- If $X$ is an absolutely continuous random variable with density $f_X$, then
  \[ E(X) = \int_{-\infty}^{\infty} x f_X(x) \, dx . \]

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

Variance

- The variance of a random variable measures the spread of the probability distribution and is the expected squared distance of the random variable from its mean, $E(X) = \mu_X$.
  \[ \text{Var}(X) = \sigma_X^2 = E((X - \mu_X)^2) . \]

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

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

Means and Variances of Important Distributions

- Discrete distributions.
  - Degenerate distributions: $P(X = c) = 1$ for some $c$.
    $E(X) = c, \text{Var}(X) = 0$.
  - Bernoulli distributions: $X \sim \text{Bernoulli}(\theta)$.
    $P(X = 1) = \theta, P(X = 0) = 1 - \theta$ where $0 < \theta < 1$ is a parameter.
    $E(X) = \theta, \text{Var}(X) = \theta(1 - \theta)$.
  - Binomial distributions: $X \sim \text{Binomial}(n, \theta)$.
    $P(X = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$ for $k = 0, 1, \ldots, n$.
    $E(X) = n\theta, \text{Var}(X) = n\theta(1 - \theta)$.
  - Geometric distributions: $X \sim \text{Geometric}(\theta)$.
    $P(X = k) = \theta(1 - \theta)^{k-1}$ for $k = 0, 1, 2, \ldots$.
    $E(X) = (1 - \theta)/\theta, \text{Var}(X) = (1 - \theta)/\theta^2$. 


– **Negative Binomial distributions:**

\[ X \sim \text{Negative Binomial}(r, \theta). \]

\[ P(X = k) = \binom{k + r - 1}{k} \theta^r (1 - \theta)^k \text{ for } k = 0, 1, 2, \ldots \]

\[ E(X) = r/(1 - \theta), \ Var(X) = r/(1 - \theta)^2. \]

– **Poisson distributions:**

\[ X \sim \text{Poisson}(\lambda). \text{ For parameter } \lambda > 0, \]

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

\[ E(X) = \lambda, \ Var(X) = \lambda. \]

– **Hypergeometric distributions:**

\[ X \sim \text{Hypergeometric}(N, M, n). \]

\[ P(X = k) = \binom{M}{k} \binom{N - M}{n - k} \binom{N}{n}, \]

where \( k = \max\{0, n - N + M\}, \ldots, \min\{M, n\} \).

\[ E(X) = nM/N, \ Var(X) = n(M/N)((N - M)/N)((N - n)/(N - 1)). \]

– **Uniform distributions:**

\[ X \sim \text{Uniform}(L, R). \]

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

\[ E(X) = (L + R)/2, \ Var(X) = (R - L)^2/12. \]

– **Exponential distributions:**

\[ X \sim \text{Exponential}(\lambda). \text{ For parameter } \lambda > 0, \]

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

\[ E(X) = 1/\lambda, \ Var(X) = 1/\lambda^2. \]

– **Gamma distributions:**

\[ X \sim \text{Gamma}(\alpha, \lambda), \text{ for shape parameter } \alpha > 0 \text{ and scale parameter } \lambda > 0. \]

The density is

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

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

is the gamma function, defined when \( \alpha > 0. \)

\[ E(X) = \alpha/\lambda, \ Var(X) = \alpha/\lambda^2. \]

– **Normal distributions:**

\[ X \sim \text{Normal}(\mu, \sigma^2). \]

The density is

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)/\sigma^2} \text{ for } -\infty < x < \infty. \]

\[ E(X) = \mu, \ Var(X) = \sigma^2. \]

– **Beta distributions:**

\[ X \sim \text{Beta}(a, b). \text{ With parameters } a > 0, b > 0, \]

the density is

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

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

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

\[ E(X) = a/(a + b), \ Var(X) = ab/((a + b)^2(a + b + 1)). \]

**Covariance**

- The covariance of random variables \( X \) and \( Y \) is

\[ \text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) \]

where \( \mu_X = E(X) \) and \( \mu_Y = E(Y). \)

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

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

\[ \text{Cov} \left( \sum_i a_i X_i, \sum_j b_j Y_j \right) = \sum_i \sum_j a_i b_j \text{Cov}(X_i, Y_j). \]

- **(Variance of Sums)**

\[ \text{Var} \left( \sum_i X_i \right) = \sum_i \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \]

- The correlation between random variables \( X \) and \( Y \) is

\[ \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\text{SD}(X)\text{SD}(Y)}, \]

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

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

**Probability Generating Functions**

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

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

\[ P(X = k) = r^{(k)}(0)/k! \]

- **(Uniqueness)** If two discrete random variables whose possible values are all nonnegative integers have the same probability generating function which is finite for some \( t_a > 0 \), 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 $m(t) = E(e^{tX})$ is the moment generating function of $X$.
- The $k$th moment of $X$ is the expected value of the $k$th power of $X$, $E(X^k)$.
- 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 $k$th derivative evaluated at 0 is equal to the $k$th moment of $X$.

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

- (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 $c(t) = E(e^{itX})$ 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 $k$th derivative of $c(t)$ evaluated at 0 is related to the $k$th moment of $X$.

$$c^{(k)}(0) = i^kE(X^k)$$

- (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 $A$,

$$E(X \mid A) = \sum_{x \in \mathbb{R}} xP(X = x \mid A) \quad \text{if } X \text{ is discrete}$$

$$= \int_{-\infty}^{\infty} x f_{X\mid A}(x) \quad \text{if } X \text{ is absolutely continuous}$$

For random variables $X$ and $Y$,

$$E(X \mid Y = y) = \sum_{x \in \mathbb{R}} xP(X = x \mid Y = y) \quad \text{if } X \text{ is discrete}$$

$$= \int_{-\infty}^{\infty} x f_{X\mid Y}(x \mid y) \quad \text{if } X \text{ is absolutely continuous}$$

- The conditional expectation $E(X \mid Y)$ is a function of $Y$, and is thus a random variable.
- (The Law of TotalExpectation) For any two random variables $X$ and $Y$,

$$E(X) = E(E(X \mid Y)) .$$

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

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

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

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

Inequalities

- 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.

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

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

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

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

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

- 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

$$g(E(X)) \leq E(g(X)).$$