Expectation: Outline
Chapter 4

- Definitions and Examples
- Properties:
- Covariance and Correlation
- Moment Generating Functions
- Sums
Expectation: Definitions

Review: The Discrete Case. If $X$ is discrete with PMF $f$, then

$$E(X) = \sum_{x \in \mathcal{X}} xf(x),$$  \hspace{1cm} (1)

where $\mathcal{X}$ is the range of $X$, provided that the sum converges absolutely.

The Continuous Case. If $X$ has density $f$, then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx,$$  \hspace{1cm} (2)

provided that the integral converges absolutely.

Shorthand Notation: Combine (1) and (2) by writing,

$$E(X) = \int_{-\infty}^{\infty} xdF(x).$$

Note: Riemann-Stieltjes Integral.
Examples

Review

Exponential Distributions. If $X \sim \text{Exp}(\lambda)$, then $f(x) = \lambda e^{-\lambda x}$ for $0 \leq x < \infty$ and

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} \, dx = \frac{1}{\lambda}$$

Poisson Distributions. If $X \sim \text{Poisson}(\lambda)$, then

$$f(x) = \frac{1}{x!} \lambda^x e^{-\lambda}$$

for $x = 0, 1, 2 \cdots$, and

$$E(X) = \sum_{x=0}^\infty x \frac{1}{x!} \lambda^x e^{-\lambda} = \lambda \sum_{x=1}^\infty \frac{1}{(x-1)!} \lambda^{x-1} e^{-\lambda} = \lambda.$$
Transformations

Suppose

\[ X_1, \ldots, X_m \sim f \] joint density or PMF.

Let

\[ Y = w(X_1, \ldots, X_m). \]

**Theorem.** If \( f \) is a PMF, then

\[ E(Y) = \sum_{x \in \mathcal{X}} w(x)f(x), \]

provided that the sum converges absolutely. If \( f \) is a density, then

\[ E(Y) = \int_{\mathbb{R}^m} w(x)f(x)dx, \]

provided that the integral converges absolutely.

**Note:** By definition, \( E(Y) = \int_{-\infty}^{\infty} ydF_Y(y) \).
An Example
Distance Between Two Points

Suppose

\[ X, Y \sim^{\text{ind}} \text{Unif}[0, 1]. \]

Then

\[ f(x, y) = \begin{cases} 
1 & \text{if } 0 \leq x, y \leq 1 \\
0 & \text{if otherwise} 
\end{cases} \]

Let

\[ D = |Y - X|. \]

Then

\[ E(D) = \int_0^1 \left[ \int_0^1 |y - x| \, dy \right] \, dx = \cdots = \frac{1}{3}. \]
**The Calculation:** Here

\[\int_0^1 |y - x| \, dy = \int_0^x (x - y) \, dy + \int_x^1 (y - x) \, dy\]

\[= -\frac{1}{2} (y - x)^2 \bigg|_{y=0}^x + \frac{1}{2} (y - x)^2 \bigg|_{y=x}^1 \]

\[= \frac{1}{2} [x^2 + (1 - x)^2].\]

So,

\[E(D) = \frac{1}{2} \int_0^1 [x^2 + (1 - x)^2] \, dx\]

\[= \frac{1}{6} [x^3 - (1 - x)^3] \bigg|_{x=0}^1\]

\[= \frac{1}{3}.\]
Properties of Expectation

**Linearity:** If $X$ and $Y$ are JDRVs with finite expectations and if $a, b \in IR$, then

$$E(aX + bY) = aE(X) + bE(Y).$$

More generally, if $X_1, \cdots, X_n$ are JDRVs with expectations and $c_1, \cdots, c_n \in IR$, then

$$E(c_1X_1 + \cdots + c_nX_n) = c_1E(X_1) + \cdots + c_nE(X_n).$$

**Counting Variables:** If $A_1, \cdots, A_n$ are events and $X = 1_{A_1} + \cdots + 1_{A_n}$, then $E(1_{A_i} = p$ and

$$E(X) = P(A_1) + \cdots + P(A_n).$$

**Example:** Binomial. If $P(A_i) = p$, then $E(X) = np$. 
More Properties of Expectation

**Monotonicity:** If $E(X)$ of $E(Y)$ are defined and $P[X \leq Y] = 1$, then

\[ E(X) \leq E(Y). \]

**Expectation and Independence:** If $X$ and $Y$ are independent RVs with expectations, then

\[ E(XY) = E(X)E(Y). \]
**Covariance and Correlation**

**Def:** If $X$ and $Y$ are JDRVs with finite means and variances, then

$$
\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]
$$

is called the covariance between $X$ and $Y$; and

$$
\rho = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}
$$

is called correlation between $X$ and $Y$

**Alternative Expression:**

$$
\sigma_{XY} = E(XY) - \mu_X \mu_Y.
$$

**Special Cases:**

a) $\sigma_{XX} = \sigma_X^2$.

b) If $X$ and $Y$ are independent, then $\sigma_{XY} = 0$.

**Note:** Measures of dependence.
Example

If

\[ X \sim \text{Unif}[-1, 1], \]
\[ Y = X^2, \]

then

\[ E(X) = \frac{1}{2} \int_{-1}^{1} x \, dx = 0, \]
\[ E(XY) = E(X^3) = \frac{1}{2} \int_{-1}^{1} x^3 \, dx = 0. \]

So,

\[ \sigma_{XY} = E(XY) - E(X)E(Y) = 0 - 0 = 0, \]

but \( X \) and \( Y \) are not independent.
Linear Functions

If

\[ X' = aX + b \quad \text{and} \quad Y' = cY + d, \]

then

\[ E(X') = aE(X) + b, \quad E(Y') = cE(Y) + d, \]

\[ \sigma_{X', Y'} = E[(X' - \mu_X')(Y' - \mu_Y')] = E[ac(X - \mu_X)(Y - \mu_Y)] = ac\sigma_{X, Y}, \]

Thus,

\[ \sigma^2_{X'} = a^2\sigma^2_X, \quad \sigma^2_{Y'} = c^2\sigma^2_Y, \]

and

\[ \rho_{X'Y'} = \frac{\sigma_{X'Y'}}{\sigma_{X'}\sigma_{Y'}} = \frac{ac\sigma_{XY}}{|a|\sigma_X|c|\sigma_Y} = \frac{ac}{|ac|}\rho_{XY}, \]

if all defined. So,

\[ |\rho_{X'Y'}| = |\rho_{XY}|, \]

and \(|\rho|\) is invariant under linear transformations.
**The Variance of Sum**

**Sums**: If $X_1, \cdots, X_n$ are jointly distributed random variables with finite variances and $S = X_1 + \cdots + X_n$, then

$$E(S) = E(X_1) + \cdots + E(X_n)$$

and

$$\sigma_S^2 = \sum_{i=1}^{m} \sigma_{X_i}^2 + 2 \sum_{j=2}^{m} \sum_{i=1}^{j-1} \sigma_{X_i, X_j}.$$ 

**Def**: $X_1, \cdots, X_m$ are **uncorrelated** if $\sigma_{X_i, X_j} = 0$ for all $i \neq j$.

**Corollary**: If $X_1, \cdots, X_m$ are uncorrelated, then

$$\sigma_S^2 = \sum_{i=1}^{m} \sigma_{X_i}^2.$$  

(*)&

In particular, (*) holds if $X_1, \cdots, X_m$ are independent.
**Example: Counting Variables.** If $A_1, \cdots, A_n$ are independent and $P(A_i) = p$ for all $i$, then $S = 1_{A_1} + \cdots + 1_{A_1} \sim \text{Binomial}(n, p)$. In this case

$$E(1_{A_i}) = p \quad \text{and} \quad E(1_{A_i}^2) = E(1_{A_i}) = p.$$ 

So, the variance of $1_{A_i}$ is $p - p^2 = pq$, $E(S) = p + \cdots + p = np$, and

$$\sigma^2_S = pq + \cdots + pq = npq.$$
Moment Generating Functions

If \( X \sim F \), then

\[
M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} dF(x)
\]

is called the moment generating function of \( X \) and/or \( F \), provided that it converges for all \( t \) in some non-degenerate interval.

**Example**: Exponential. If \( X \sim \text{Exp}(\lambda) \), then

\[
M(t) = \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} \, dx
\]

\[
= \lambda \int_{0}^{\infty} e^{-(\lambda-t)x} \, dx
\]

\[
= -\lambda \left. \frac{e^{-(\lambda-t)x}}{\lambda - t} \right|_{x=0}^{\infty} = \frac{\lambda}{\lambda - t},
\]

for \( t < \lambda \).
Other Examples

See The Text

**Poisson.** If

\[ X \sim \text{Poisson}(\lambda), \]

then

\[ M(t) = e^{\lambda(e^t-1)} \]

for all \(-\infty < t < \infty\).

**Normal.** If

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

then

\[ M(t) = e^{\mu t + \frac{1}{2} \sigma^2 t^2} \]

for \(-\infty < t < \infty\).
Moments and the MGF

Moments: Recall

\[ \mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k dF(x). \]

Moments and Derivatives. If \( M(t) < \infty \) for \( |t| < h \) for some \( h > 0 \), then

\[ M(0) = 1, \]
\[ M'(0) = \mu, \]
\[ M''(0) = \mu_2, \]
\[ \ldots \]
\[ M^{(k)}(0) = \mu_k, \]

for all \( k = 0, 1, 2, \ldots \).
Proof-Outline-The Discrete Case. If $X$ has PMF $f$, then

$$M(t) = \sum_{x \in \mathcal{X}} e^{tx} f(x),$$

$$M(0) = \sum_{x \in \mathcal{X}} 1 \times f(x) = 1,$$

$$M'(t) = \sum_{x \in \mathcal{X}} xe^{tx} f(x),$$

$$M'(0) = \sum_{x \in \mathcal{X}} xf(x) = \mu,$$

...
Example
Exponential

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

$$M(t) = \frac{\lambda}{\lambda - t},$$

$$M'(t) = \frac{\lambda}{(\lambda - t)^2},$$

$$M''(t) = \frac{2\lambda}{(\lambda - t)^3},$$

$$\cdots,$$

$$M^{(k)}(t) = \frac{k!\lambda}{(\lambda - t)^{k+1}}.$$

So,

$$\mu_k = \frac{k!}{\lambda^k}$$

for $k = 1, 2, \cdots$. 
Alternative Formulation

Let

\[ m(t) = \log[M(t)]. \]

Then

\[ m'(t) = \frac{M'(t)}{M(t)}, \]

and

\[ m''(t) = \frac{M(t)M''(t) - M'(t)^2}{M(t)^2}. \]

So,

\[ m'(0) = \frac{M'(0)}{M(0)} = \mu \]

and

\[ m''(0) = \mu_2 - \mu^2 = \sigma^2. \]

**Cumulants:** \( \kappa_j = m^{(j)}(0). \)
Sums

If $X_1, \cdots, X_n$ are independent with MGFs $M_1, \cdots, M_n$, then the MGF of

$$S = X_1 + \cdots + X_n$$

is

$$M_S(t) = M_1(t) \times \cdots \times M_n(t).$$

Proof. We have

$$M_S(t) = E(e^{tS}) = E\left( \prod_{i=1}^{n} e^{tX_i} \right) = \prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n} M_i(t).$$

Note: Product, not convolution.
**Unicity:** If $M_X(t) = M_Y(t)$ for all $t$ in some non-degenerate interval, then $F_X(z) = F_Y(z)$ for all $z$.

**Example: Normal:** If $X_1, \cdots, X_n$ are independent and

$$X_i \sim^{	ext{ind}} \text{Normal}(\mu_i, \sigma_i^2),$$

then

$$S \sim \text{Normal}(\mu, \sigma^2),$$

where $\mu = \mu_1 + \cdots + \mu_n$ and $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$, since

$$M_S(t) = \prod_{i=1}^{n} M_i(t) = \prod_{i=1}^{n} e^{\mu_i t + \frac{1}{2} \sigma_i^2 t^2} = e^{\mu t + \frac{1}{2} \sigma^2 t^2}$$

and, therefore, $S \sim \text{Normal}(\mu, \sigma^2)$. 
Limit Theorems
Outline

- Inequalities
- Convergence in Probability
- The Law of Large Numbers
- The Central Limit Theorem
Inequalities

Markov’s. If $Y$ is any RV and $0 < c < \infty$, then

$$P[|Y| \geq c] \leq \frac{1}{c} E|Y|.$$ 

Proof. Let $B = \{|Y| \geq c\}$. Then $c1_B \leq |Y|$. So,

$$E|Y| \geq E(c1_B) = cP(B).$$

Chebyshev’s Inequality: If $X$ has mean $\mu = E(X)$ and variance $\sigma^2 = E[(X - \mu)^2]$, then

$$P[|X - \mu| \geq c] \leq \frac{\sigma^2}{c^2}.$$ 

Proof. Let $Y = |X - \mu|^2$ in Markov’s Inequality. Then

$$P[|X - \mu| \geq c] = P[Y \geq c^2] \leq \frac{1}{c^2} E|Y| = \frac{\sigma^2}{c^2}.$$
Sums of Independent RVs

If $X_1, \cdots, X_n$ and are independent, with means $\mu_i$ and variances $\sigma_i^2$, and $S = X_1 + \cdots + X_n$, then

$$E(S) = \mu_1 + \cdots + \mu_n,$$

$$\sigma_S^2 = \sigma_1^2 + \cdots + \sigma_n^2.$$ 

**Special Case:** If $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$ for $i = 1, \cdots, n$, then

$$E(S) = n\mu, \quad \sigma_S^2 = n\sigma^2,$$

and

$$\sigma_S = \sigma \sqrt{n}.$$ 

**Note:** $E(S)$ grow like $n$; $\sigma_S$ grows like $\sqrt{n}$. 
Type of Convergence

If $Y_n$ are RVs and $c$ is a constant, then $Y_n$ converges to $c$ in mean square iff

$$\lim_{n \to \infty} E[(Y_n - c)^2] = 0;$$

and $Y_n$ converges to $c$ in probability iff

$$\lim_{n \to \infty} P[|Y_n - c| \geq \epsilon] = 0$$

for all $\epsilon > 0$.

A Simple Relation: If $Y_n \to^{ms} c$, then $Y_n \to^{p} c$, since

$$P[|Y_n - c| \geq \epsilon] = P[|Y_n - c|^2 \geq \epsilon^2]$$

$$\leq \frac{1}{\epsilon^2} E|Y_n - c|^2,$$

by Markov’s Inequality.
**The Law of Large Numbers**

**Theorem.** Let $F$ be a distribution function with mean and variance

$$
\mu = \int_{-\infty}^{\infty} x dF(x) \quad \text{and} \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x);
$$

and let $X_1, \cdots, X_n \sim^{ind} F$. Then

$$
\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \to \mu
$$

in mean square and in probability.

**Proof.** Here

$$
E[(\bar{X}_n - \mu)^2] = \sigma^2_{\bar{X}_n} = \frac{\sigma^2}{n} \to 0,
$$

**Paraphrase:** Time Average $= $ Space Average
Indicators

If $A_1, A_2, \cdots$ are independent with $P(A_i) = p$, then $E(1_{A_i}) = p$, and

$$\frac{1}{n}[1_{A_1} + \cdots + 1_{A_n}] \to p.$$  

**Note:** This is the frequentist interpretation of “probability.”

**The Law of Averages:** Let $N_n = 1_{A_1} + \cdots + 1_{A_n}$. Then

$$P[A_{n+1}\mid N_n = k] = p$$

for all $k$ (assuming that $p$ is fixed).
Normal Approximation

Recall: If \( X_1, X_2, \ldots \) are independent with finite means \( \mu_i \) and variances \( \sigma_i^2 \), then the mean and variance of \( S_n = X_1 + \cdots + X_n \) are

\[
E(S_n) = \mu_1 + \cdots + \mu_n \quad \text{and} \quad \sigma^2_{S_n} = \sigma_1^2 + \cdots + \sigma_n^2.
\]

Let

\[
S^*_n = \frac{S_n - E(S_n)}{\sigma_{S_n}}
\]

The Central Limit Theorem. For all real \( z \),

\[
\lim_{n \to \infty} P[S^*_n \leq z] = \Phi(z),
\]

provided that each \( X_i \) contributes negligibly to the sum. In particular, (*) holds if \( X_1, \cdots, X_n \) are i.i.d.
Example

Let $T = \text{Tax}$, $\langle T \rangle = \text{closest integer}$, and $X = T - \langle T \rangle$. Suppose

$$X \sim \text{Unif}(\frac{-1}{2}, \frac{1}{2}).$$

Then

$$\mu = E(X) = 0 \quad \text{and} \quad \sigma^2 = Var(X) = \frac{1}{12}.$$ 

Let $n = 12,000,000$, $X_1, \cdots, X_n \sim^{ind} \text{Unif}(\frac{-1}{2}, \frac{1}{2})$, and

$S = X_1 + \cdots + X_n$.

**Question:** How big is $S$?

**Worst Case Analysis:** $|S| \leq 6,000,000$, but this is very conservative.
**Probabilistic Analysis:** By CLThm,

\[ S \approx \text{Normal}[n\mu = 0, n\sigma^2 = 1,000,000]. \]

**Note:**

\[ \sqrt{n\sigma^2} = 1000. \]

So,

\[
P[-2000 \leq S \leq 2000] \\
\approx \Phi\left(\frac{2000 - 0}{1000}\right) - \Phi\left(\frac{-2000 - 0}{1000}\right) \\
= \Phi(2) - \Phi(-2) \\
= .954.
\]