1 Conditional Expectation

**Definition 1.1** (Total variance). \( \text{var}(X|Y) := \mathbb{E}[(X - \mathbb{E}[X|Y])^2 | Y] = \mathbb{E}[X^2 | Y] - (\mathbb{E}[X|Y])^2. \)

**Theorem 1.2** (Law of total variance). \( \text{var}(X) = \mathbb{E}(\text{var}(X|Y)) + \text{var}(\mathbb{E}(X|Y)). \)

**Exercise 1.** Let \( n \in \mathbb{N} \) be given. Suppose that
- \( U \sim \text{Uniform}(0,1) \), and
- the distribution of \( X \) given \( U = p \) is Binomial\((n,p)\).

Find the distribution of \( X \). (*Hint. We have \( \int_0^1 x^a(1-x)^b \, dx = a!b!/(a+b+1)! \) for \( a,b \geq 0 \).*)

**Solution.** (1) Range of \( X \) is \( \{0,1,\cdots,n\} \). (2) For each \( k = 0,1,\cdots,n \), we have

\[
\mathbb{P}(X = k) = \int_{-\infty}^{\infty} \mathbb{P}(X = k|U = p) \, f_U(p) \, dp = \int_0^1 \mathbb{P}(X = k|U = p) \, dp
\]

\[
= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \, dp = \frac{1}{n+1}.
\]

Therefore \( X \) has the (discrete) uniform distribution on \( \{0,1,\cdots,n\} \). \( \square \)

**Exercise 2.** Let \( U_1, U_2, \cdots \) be i.i.d. and \( U_i \sim \text{Uniform}(0,1) \). For \( x > 0 \) define

\[
N(x) = \min\{n : x < U_1 + \cdots + U_n\} \quad \text{and} \quad m(x) = \mathbb{E}N(x).
\]

Show that \( m(x) \) satisfies

\[
m(x) = 1 + \int_0^x m(u) \, du \quad \text{for} \quad 0 < x \leq 1.
\] (1.1)

(It is well known that (1.1) has a unique solution \( m(x) = e^x \), so we have \( EN(1) = e \).)

**Solution.** We use the law of iterated expectation. For \( 0 < x \leq 1 \),

\[
m(x) = \mathbb{E}N(x) = \mathbb{E}[\mathbb{E}[N(x)|U_1]] = \int_0^1 \mathbb{E}[N(x)|U_1 = u] \, du.
\]

But we observe that
• If \( u > x \), then \( N(x) = 1 \).
• If \( u \leq x \), then
  \[ N(x) = 1 + \min\{n : U_2 + \cdots + U_{n+1} > x - u\}. \]

Also, if we put \( \tilde{N}(x) = \min\{n : U_2 + \cdots + U_{n+1} > x\} \), then \( \tilde{N}(x) \) and \( N(x) \) has the same distribution and hence \( \mathbb{E} \tilde{N}(x) = m(x) \).

So it follows that
\[
m(x) = \int_0^x \mathbb{E}[N(x)|U_1 = u] \, du + \int_x^1 \mathbb{E}[N(x)|U_1 = u] \, du
= \int_0^x (1 + \mathbb{E} \tilde{N}(x-u)) \, du + \int_x^1 1 \, du
= 1 + \int_0^x m(x-u) \, du = 1 + \int_0^x m(u) \, du.
\]

This proves the claim. \[\square\]

Exercise 3. You play a game as follows:
• If you start the game, you are given a random positive integer \( N \).
• Given \( N = n \), toss a biased coin \( n \) (of getting head with probability \( p \)) times independently, and record the total number of heads by \( X \).

Find the expectation \( \mathbb{E}X \) and the variance \( \text{var}(X) \) in terms of \( \mathbb{E}N \) and \( \text{var}(N) \).

Solution.
• **Expectation.** The distribution of \( X \) given \( N = n \) is Binomial\((n,p)\). So \( \mathbb{E}[X|N = n] = np \). Then By the law of iterated expectation,
  \[ \mathbb{E}X = \mathbb{E}[\mathbb{E}[X|N]] = \mathbb{E}[Np] = \mathbb{E}[N]p. \]

• **Variance.** From the law of total variance,
  \[ \text{var}(X) = \mathbb{E}[\text{var}(X|N)] + \text{var}(\mathbb{E}[X|N]) = \mathbb{E}[Np(1 - p)] + \text{var}(Np) = \mathbb{E}[N]p(1 - p) + \text{var}(N)p^2. \]

\[\square\]

2 Transforms

**Definition** (Transform). \( M_X(s) := \mathbb{E}[e^{sX}] \).

**Example 2.1.** The followings are some examples:
• If \( X \) is discrete, then \( M_X(s) = \sum_x e^{sx} p_X(x) \).
• If \( X \sim \text{Poisson}(\lambda) \), then \( M_X(s) = e^{(e^s-1)} \).
• If \( X \sim \text{Exp}(\lambda) \), then \( M_X(s) = \frac{1}{\lambda - s} \) for \( s < \lambda \) and \( M_X(s) = +\infty \) for \( s \geq \lambda \).
• If \( X \sim \text{N}(0,1) \), then \( M_X(s) = e^{s^2/2} \).
Theorem 2.2. We have \( \frac{d^n}{ds^n} M_X(s) = \mathbb{E}[X^n e^{sX}] \). In particular, \( M_X^{(n)}(0) = \mathbb{E}[X^n] \).

Theorem 2.3 (Inversion property). \( M_X(s) \) uniquely determines the distribution (i.e. CDF) of \( X \) whenever \( M_X(s) \) is finite near \( s = 0 \).

Proposition 2.4. The followings hold:
- \( M_{aX+b}(s) = e^{bs} M_X(as) \).
- If \( X \) and \( Y \) are independent, \( M_{X+Y}(s) = M_X(s)M_Y(s) \).

Exercise 1. The transform of a random variable \( X \) is
\[
M_X(s) = \left( \frac{1}{2} + \frac{e^s}{2} \right) e^{e^{-s} - 1}.
\]
Compute the probability \( P(X = 0) \).

Solution. Notice that
\[
M_X(s) = \left( \frac{1}{2} + \frac{e^s}{2} \right) \frac{1}{e} \sum_{n=0}^{\infty} \frac{e^{ns} n!}{n!} = \frac{1}{2} e \sum_{n=0}^{\infty} \frac{e^{ns}}{n!} + \frac{1}{2} e \sum_{n=0}^{\infty} \frac{e^{(n+1)s}}{n!}
\]
Comparing the general formula, we find that \( X \) is a discrete random variable
\[
P(X = 0) = \frac{1}{2e} \quad \text{and} \quad P(X = n) = \frac{n + 1}{2e(n!)} \quad \text{for} \quad n \geq 1.
\]

Exercise 2. The transform of a random variable \( X \) is
\[
M_X(s) = \frac{2 - s}{2(1 - s)^2}.
\]
Compute \( \mathbb{E}X \).

Solution. Differentiating \( M_X(s) \), we get
\[
\frac{d}{ds} M_X(s) = \frac{3 - s}{2(1 - s)^3}
\]
Plugging \( s = 0 \) gives
\[
\mathbb{E}X = \left. \frac{d}{ds} M_X(s) \right|_{s=0} = \frac{3}{2}
\]
Alternatively, you can check the following function
\[
f_X(x) = \begin{cases} \frac{1}{2}(x + 1)e^{-x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}
\]
is indeed the PDF of \( X \). Then you can calculate \( \mathbb{E}X \) directly as follows:
\[
\mathbb{E}X = \int_{0}^{\infty} x f_X(x) \, dx = \frac{1}{2} \int_{0}^{\infty} x(x + 1) e^{-x} \, dx = \frac{3}{2}.
\]
3 (optional) Sample mean and sample variance

Exercise 1. Let $X_1, \cdots, X_n$ be i.i.d random variables. Let $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.

(a) Show that $\mathbb{E}[g(X_i)|\bar{X}] = \mathbb{E}[g(X_i)|\bar{X}]$ regardless of the choice of $g$.
(b) Show that $\mathbb{E}[X_i|\bar{X}] = \bar{X}$.

In statistics, $\bar{X}$ is called the sample mean or empirical mean.

Solution.
(a) Notice the following two observations:
- $\bar{X} = \bar{x}$ is equivalent to $X_1 + \cdots + X_n = n\bar{x}$.
- Interchanging the role of $X_1$ and $X_2$ does not change the condition $\bar{X} = \bar{x}$.

So it follows that
\[
\mathbb{E}[g(X_i)|\bar{X} = \bar{x}] = \mathbb{E}[g(X_i)|X_1 + X_2 + \cdots + X_n = n\bar{x}]
= \mathbb{E}[g(X_i)|X_2 + X_1 + \cdots + X_n = n\bar{x}]
= \mathbb{E}[g(X_2)|\bar{X} = \bar{x}]
\]
and hence $\mathbb{E}[g(X_i)|\bar{X}] = \mathbb{E}[g(X_2)|\bar{X}]$. The general case follows likewise.
(b) From the previous part with $g(x) = x$,
\[
\mathbb{E}(X_i|\bar{X}) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_i|\bar{X}) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_k|\bar{X}) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} X_k|\bar{X}\right) = \mathbb{E}([\bar{X}|\bar{X}] = \bar{X}.
\]

Remark. In other words, only knowing the mean $\bar{X}$ of $n$ samples, the best information you can hope for $X_i$ is only the mean $\bar{X}$ itself.

Exercise 2. Let $n \geq 2$ and $X_1, \cdots, X_n$ be i.i.d random variables. Also let $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.

(a) Show that $X_i - \bar{X}$ and $\bar{X}$ are uncorrelated.
(b) Let $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Show that $\mathbb{E}[S^2|\bar{X}] = \frac{n}{n-1} \text{var}(X_1|\bar{X})$.
(c) Show that $\mathbb{E}(S^2) = \text{var}(X_1)$.

In statistics, $S^2$ is called the (unbiased) sample variance.

Solution. (a) Notice that $\mathbb{E}[X_i] = \mathbb{E}X_i$ for any $i$. This shows that
\[
\text{cov}(X_i - \bar{X}, \bar{X}) = \mathbb{E}[(X_i - \bar{X})\bar{X}] - \mathbb{E}[X_i - \bar{X}] \mathbb{E}[\bar{X}] = \mathbb{E}[(X_i - \bar{X})\bar{X}] = 0.
\]
On the other hand, by the previous example, we have
\[
\mathbb{E}[(X_i - \bar{X})\bar{X}|\bar{X}] = \mathbb{E}[X_i - \bar{X}|\bar{X}]\bar{X} = (\bar{X} - \bar{X})\bar{X} = 0.
\]
Now by the law of iterated expectation and this observation, we get
\[
\mathbb{E}[(X_i - \bar{X})\bar{X}] = \mathbb{E}[\mathbb{E}[(X_i - \bar{X})\bar{X}|\bar{X}]] = 0.
\]
(b) We have
\[
\mathbb{E}[S^2|\bar{X}] = \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E}[(X_i - \bar{X})^2]\bar{X}.
\]

(*)
But we know that $\bar{X} = \mathbb{E}[X_i | \bar{X}]$. Consequently

$$\mathbb{E}[(X_i - \bar{X})^2 | \bar{X}] = \mathbb{E}[(X_i - \mathbb{E}[X_i | \bar{X}])^2 | \bar{X}] = \text{var}(X_i | \bar{X})$$

and also

$$\text{var}(X_i | \bar{X}) = \mathbb{E}[X_i^2 | \bar{X}] - (\mathbb{E}[X_i | \bar{X}])^2 = \mathbb{E}[X_i^2 | \bar{X}] - (\mathbb{E}[X_i | \bar{X}])^2 = \text{var}(X_i | \bar{X}).$$

Therefore (*) becomes

$$\mathbb{E}[S^2 | \bar{X}] = \frac{1}{n-1} \sum_{i=1}^{n} \text{var}(X_i | \bar{X}) = \frac{n}{n-1} \text{var}(X_1 | \bar{X}).$$

(c) From the law of iterated expectation,

$$\mathbb{E}[S^2] = \mathbb{E}[\mathbb{E}[S^2 | \bar{X}]] = \frac{n}{n-1} \mathbb{E}[\text{var}(X_1 | \bar{X})]. \quad (*)$$

Then by the law of total variation,

$$\mathbb{E}[S^2] = \frac{n}{n-1} (\text{var}(X_1) - \text{var}(\mathbb{E}(X_1 | \bar{X}))) = \frac{n}{n-1} (\text{var}(X_1) - \text{var}(\bar{X}))$$

$$= \frac{n}{n-1} \left( \text{var}(X_1) - \frac{1}{n} \text{var}(X_1) \right) = \text{var}(X_1).$$

□