Feb 1-2016

$X$: a r.v.

$E[g(X)] = \begin{cases} \sum_{x \in X} g(x)f_X(x) & \text{if } X \text{ is discrete} \\ \int g(x)f_X(x)dx & \text{if } X \text{ is continuous} \end{cases}$

Properties

1. $E[c] = c$
2. If $g(x) \geq 0 \text{ w.p. } 1$, then $E[g(x)] \geq 0$
3. Linearity
   
   $E[\alpha g_1(x) + g_2(x)] = \alpha E[g_1(x)] + E[g_2(x)]$
4. Fundamental theorem of expectation
   
   If $Y = g(X)$, then
   
   $E[Y] = E[g(X)]$

- Computed from the distribution of $X$.

Mean and variance

- The first moment (mean) of $X$ is $E(X) = \frac{\int x f_X(x)dx}{\sum x f_X(x)}$

**Trick:** If $X \geq 0$, then $E[X] = \int _0^\infty (1 - F_X(x))dx$

Another trick (indicator variable)

$1_A = \begin{cases} 1 & \text{if event } A \text{ happens} \\ 0 & \text{o.w.} \end{cases}$

$E[1_A] = P(A). \quad (\int 1_A f(x)dx = \int_A f(x)dx = P(x \in A))$

- The second moment of $X$ is $E(X^2) = \left\{ \int x^2 f_X(x)dx \right\} \sum x^2 p_X(x)$

- The variance of $X$ is $\text{Var}(X) = E[(X - E(X))^2]$ (The standard deviation of $X$ is $\sigma_X = \sqrt{\text{Var}(X)}$)

$=\int (x^2 - 2x \cdot E(X) + (E(X))^2) f_X(x)dx$

$= EX^2 - 2(EX)^2 + (EX)^2$ (i.e. $\text{Var}(X) = \sigma_X^2$)

$= EX^2 - (EX)^2$

$\text{since } \text{Var}(X) \geq 0, \quad E(X^2) \geq (EX)^2$. 

### Mean and variance of famous distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bern(p)</td>
<td>p</td>
<td>p(1-p)</td>
</tr>
<tr>
<td>Geom(p)</td>
<td>1/p</td>
<td>( \frac{1-p}{p^2} )</td>
</tr>
<tr>
<td>Bihom(n,p)</td>
<td>np</td>
<td>np(1-p)</td>
</tr>
<tr>
<td>Poisson(λ)</td>
<td>λ</td>
<td>( \frac{λ}{1-λ} )</td>
</tr>
<tr>
<td>U[a,b]</td>
<td>( \frac{a+b}{2} )</td>
<td>( \frac{(b-a)^2}{12} )</td>
</tr>
<tr>
<td>Exp(λ)</td>
<td>( \frac{1}{λ} )</td>
<td>( \frac{1}{λ^2} )</td>
</tr>
<tr>
<td>N(μ,σ²)</td>
<td>μ</td>
<td>σ²</td>
</tr>
</tbody>
</table>

**Note**

1. **Expectation can be infinite**
   
   For example, if \( f_X(x) = \begin{cases} \frac{1}{x^2}, & x \geq 1 \\ 0, & \text{o.w.} \end{cases} \)
   
   \[ E(X) = \int_1^{\infty} x \frac{1}{x^2} \, dx = \int_1^{\infty} \frac{1}{x} \, dx = \ln x \bigg|_1^{\infty} = \infty. \]

2. **Expectation may not exist**
   For example, the Cauchy distribution has the pdf \( f_X(x) = \frac{1}{\pi(1+x^2)} \)
   
   \[ E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)} \, dx \]
   
   \[ = \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)} \, dx + \int_{0}^{\infty} \frac{x}{\pi(1+x^2)} \, dx \]
   
   \[ = (\pm \infty) + \infty \quad = \text{(undefined)}. \]
Inequalities

(1) Markov inequality

If \( X \geq 0 \) with \( \mathbb{E}(X) = \mu < \infty \), then
\[
\Pr(X \geq \alpha \mu) \leq \frac{1}{\alpha}. \quad \alpha \geq 1
\]

Proof: Let \( A = \{ X \geq \alpha \mu \} \) and consider \( 1_A \).

Now, note that \( 1_A \leq \frac{X}{\alpha \mu} \). Taking \( \mathbb{E}[ \cdot ] \) on both sides \( \Pr(A) = \mathbb{E}[1_A] \leq \frac{\mathbb{E}X}{\alpha \mu} = \frac{1}{\alpha} \).

(2) Chebyshev inequality

Let \( X \) be a r.v. with \( \mathbb{E}X = \mu < \infty \) and \( \mathbb{V}ar(X) = \sigma^2 < \infty \).

Then
\[
\Pr(\{ |X - \mu| \geq \alpha \sigma \}) \leq \frac{1}{\alpha^2}, \quad \alpha \geq 1
\]

Proof: \( \Pr(\{ (X - \mu)^2 \geq \alpha^2 \sigma^2 \}) = \Pr(\{ Y \geq \alpha^2 \mathbb{E}Y \}) \) where \( Y = (X - \mu)^2 \).

\[
\leq \frac{1}{\alpha^2}
\]

by Markov inequality. \( \square \)

(3) Jensen’s inequality

Def. A function \( g(x) \) is said to be convex if
\[
g(x) \leq \frac{g(b) - g(a)}{b - a} (x - a) + g(a)
\]

for all \( x \in [a, b] \) and all a, b.

- If \( g(x) \) is twice differentiable,
- then \( g(x) \) is convex iff \( g''(x) \geq 0 \).

Examples
(a) \( g(x) = ax + b \)
(b) \( g(x) = x^2 \)
(c) \( g(x) = |x|^p \), \( p \geq 1 \).
(d) \( g(x) = x \log x \) (\( x > 0 \))
(e) \( g(x) = \frac{1}{x} \) (\( x > 0 \))

Jensen’s inequality

If \( g(x) \) is convex, then \( \mathbb{E}[g(X)] \geq g(\mathbb{E}X) \)

If \( g(x) \) is concave, then \( \mathbb{E}[g(X)] \leq g(\mathbb{E}X) \).
Examples

(a) \( g(x) = x^2 \Rightarrow E(x^2) \geq (EX)^2 \)

(b) Monotonicity of norms:
\[
E[|x|^p]^{\frac{1}{p}} \geq E[|x|^r]^{\frac{1}{r}} \quad p \geq r \geq 1.
\]

Proof: Let \( g(x) = x^{\frac{r}{p}} \) convex \( x > 0 \). Then by Jensen,
\[
E[(|x|^r)^{\frac{r}{p}}] \geq (E[|x|^r])^{\frac{r}{p}}
\]
\[
\Rightarrow E[|x|^r] \geq E[|x|^r]^{\frac{r}{p}}
\]
\[
\Rightarrow E[|x|^p]^{\frac{1}{p}} \geq E[|x|^r]^{\frac{1}{r}} \quad \Box
\]

(c) Let \( g(x) = \frac{1}{x} \), \( x > 0 \).

By Jensen, \( E\left[\frac{1}{x}\right] \geq \frac{1}{E(x)} \). Thus, \( E\left[\frac{1}{x^2}\right] \geq \frac{1}{E(x^2)} \)

Expectation involving two random variables

Let \( (X,Y) \sim f_{XY}(x,y) \)

Then, the expectation of \( Z = g(X,Y) \) is \( E[Z] = E[g(X,Y)] = \int g(x,y) f_{XY}(x,y) \, dx \, dy \)

Examples

1. Correlation \( E[XY] \)

We say \( X \) and \( Y \) are orthogonal if \( E[XY] = 0 \)

2. Covariance \( \text{Cov}(X,Y) = E[(X-EX)(Y-EY)] \)
\[
= E(XY) - EX \cdot EY.
\]

\( \text{Var}(X) = \text{Cov}(X,X) \)

We say \( X \) and \( Y \) are uncorrelated if \( \text{Cov}(X,Y) = 0 \).

Correlation coefficient \( \rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \in [-1, 1] \).

If \( X \) and \( Y \) are independent, then \( X \) and \( Y \) are uncorrelated.

Proof: \( E[XY] = \iint xy f_X(x) f_Y(y) \, dx \, dy = \int x f_X(x) \, dx \int y f_Y(y) \, dy = E[X]E[Y] \). \( \Box \)

The converse is not true.

Example \( \rho_{X,Y}(x,y) \)
\[
\begin{array}{ccc}
 x & 0 & 1 \\
-1 & 1/6 & 0 & 1/6 \\
0 & 0 & 1/3 & 0 \\
1 & 1/6 & 0 & 1/6
\end{array}
\]

\( EX = EY = E[XY] \), but \( X \) and \( Y \) are dependent.