Lecture Note 4 *

Expectation (Moments)

MIT 14.30 Spring 2006

Herman Bennett

7 Expected Value

7.1 Univariate Model

Let X be a RV with pmf/pdf $f(x)$. The expected or mean value of X, denoted $E(X)$ or μ_X, is defined as:

$$\begin{aligned} E(X) &= \mu_X = \sum_{x \in X} x f(x) && \text{(discrete model)} \\ E(X) &= \mu_X = \int_{-\infty}^{\infty} x f(x) dx && \text{(continuous model)} \end{aligned} \tag{18}$$

- Intuition: central tendency (a "summary" of the distribution).
- Computation: weighted average.

If $Z = z(X)$ is a new RV defined as a function (transformation) of the RV X, then:

$$\begin{aligned} E[Z] &= E[z(X)] = \mu_Z = \sum_{x \in X} z(x) f(x) && \text{(discrete model)} \\ E[Z] &= E[z(X)] = \mu_Z = \int_{-\infty}^{\infty} z(x) f(x) dx && \text{(continuous model)} \end{aligned} \tag{19}$$

*Caution: These notes are not necessarily self-explanatory notes. They are to be used as a complement to (and not as a substitute for) the lectures.

Example 7.1. a) Find $E(X)$ and $E(X^2)$, where X is the RV that represents the outcome of rolling a die. b) Find $E(Z)$ and $E(X)$, where the pdf of the RV X is $f(x) = 2x$ if $0 < x < 1$, 0 if otherwise, and $Z = \sqrt{X}$.

- Mean vs. median.

7.2 Bivariate Model

Let (X, Y) be a random vector with joint pmf/pdf $f(x, y)$. The expected or mean value of the RV $Z = z(X, Y)$ is:

$$E(Z) = \sum_{(x,y)\in(X,Y)} z(x, y)f(x, y) \qquad\qquad E(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z(x, y)f(x, y)dxdy \tag{20}$$

- The corresponding definition for more than 2 random variables are analogous (see Multivariate Distributions at the end of Lecture Note 3).

Example 7.2. Find $E(Z)$, where $f(x,y) = 1$ if $0 < x,y < 1$, 0 if otherwise, and $Z = X^2 + Y^2$.

7.3 Properties of Expected Value

Let Y, X_1, X_2,..X_n be random variables and a, b, c, and d constants. Then,

a. $E(aX + b) = aE(X) + b$ and $E[az(X) + b] = aE[z(X)] + b$.

b. $E(aX_1 + bX_2 + ...cX_n + d) = aE(X_1) + bE(X_2) + ...cE(X_n) + d$

c. X and Y independent RVs $\longrightarrow E(XY) = E(X)E(Y)$ $(\longleftarrow ?)$

Example 7.3. Show **a** and **c**. (HOMEWORK: Show **b**.)

- $E[z(X)] \stackrel{?}{=} z(E[X])$ (Jensen's inequality)

8 Variance

The variance of a random variable X, denoted $\text{Var}(X)$ or σ_X^2, is defined as:

$$\text{Var}(X) = \sigma_X^2 = E[(X - \mu_X)^2], \quad \mu_X = E(X). \tag{21}$$

- Standard deviation, $\sigma_X = \sqrt{\sigma_X^2}$

8.1 Properties of Variance

Let X_1, X_2,..X_n be random variables, and a, b, c and d constants. Then,

a. $\text{Var}(X) = 0 \longleftrightarrow \exists\, c$ s.t. $P(X = c) = 1$ (degenerate distribution).

b. $\text{Var}(X) = E(X^2) - [E(X)]^2$.

c. $\text{Var}(aX + b) = a^2\text{Var}(X)$.

d. If X_1, X_2,..X_n are independent RVs, then

$$\text{Var}(aX_1 + bX_2 + ...cX_n + d) = a^2\text{Var}(X_1) + b^2\text{Var}(X_2) + ...c^2\text{Var}(X_n)$$

Example 8.1. Show **b** and **c**. (HOMEWORK: Show **d** with 2 RVs.)

Example 8.2. Find $\text{Var}(X)$ and $\text{Var}(Y)$, where $f(x) = 1/5$ if $x = -2, 0, 1, 3, 4$, 0 if otherwise, and $Y = 4X - 7$.

Example 8.3. Find $\text{Var}(Y)$ if $Y \sim \text{bin}(n, p)$.

9 Covariance and Correlation

Let X and Y be two random variables. The covariance of X and Y, denoted $\text{Cov}(X, Y)$, is given by:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \tag{22}$$

- Correlation of X and Y: $\text{Corr}(X, Y) = \rho(X, Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$ (standardized version of $\text{Cov}(X, Y)$).

9.1 Properties of Covariance and Correlation

Let X and Y be random variables, and a, b, c, and d constants. Then,

a. $\text{Cov}(X, X) = \text{Var}(X)$.

b. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

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

d. X and Y independent $\longrightarrow$ $\text{Cov}(X, Y) = 0$.

e. $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.

f. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

g. $\rho(X, Y) \begin{cases} > 0 & \text{"positively correlated"} \\ = 0 & \text{"uncorrelated"} \\ < 0 & \text{"negatively correlated."} \end{cases}$

h. $|\rho(X, Y)| \leq 1$.

i. $|\rho(X, Y)| = 1$ *iff* $Y = aX + b$, for $a \neq 0$.

Example 9.1. Show **c**, **d**, and **f**.

Example 9.2. Find $\text{Cov}(X, Y)$ and $\rho(X, Y)$, where $f(x, y) = 8xy$ for $0 \leq x \leq y \leq 1$, 0 if otherwise.

10 Conditional Expectation and Conditional Variance

Let (X, Y) be a random vector with conditional pmf/pdf $f(y|x)$. The conditional expectation of Y given X=x, denoted $E(Y|X = x)$, is given by:

$$E(Y|X = x) = \sum_{y \in \Re} y f(y|x) \qquad \text{and} \qquad E(Y|X = x) = \int_{-\infty}^{\infty} y f(y|x) dy \tag{23}$$

(discrete model) (continuous model)

Example 10.1. Find $E(Y|X = x)$, where $f(x, y) = e^{-y}$ for $0 \leq x \leq y \leq \infty$, 0 if otherwise.

Law of Iterated Expectation. Let (X, Y) be a random vector. Then,

$$E[E(Y|X)] = E(Y) \tag{24}$$

Example 10.2. Prove (24).

Conditional Variance Identity. For any two random variables X and Y, the variance of X can be decomposed as follows:

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]. \tag{25}$$

Example 10.3. Each year an R&D firm produces N innovations according to some random process, where $E(N) = 2$ and $\text{Var}(N) = 1$. Each innovation is a commercial success with probability 0.2 and this probability is independent of previous innovations' performance. a) If there are 5 innovations this year, what is the pmf of the number of successes and its expected value? b) What is the expected number of commercial successes before knowing the number of innovations produced? c) What is the variance of the number of commercial successes before knowing the number of innovations produced?

11 Moments and Moment Generating Function

11.1 Moment

Let X be a continuous RV. The moment $E[g(X)]$ is given by:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \longrightarrow \text{ Expectation of } g(X). \tag{26}$$

- Analogously for the discrete case.
- For example, the mean value can be characterized as a moment, where $g(X) = X$.
- The n^{th} moment of X is defined as $E[X^n]$, which implies that $g(X) = X^n$.

  -Skewness.

  -Kurtosis.

11.2 Moment Generating Function

Let X be a RV. The moment generating function of X, denoted $M_X(t)$, is defined as

$$M_X(t) = E[e^{tX}] \ , \tag{27}$$

and satisfies the following property:

$$M_X^{(n)}(t) = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} = E[X^n] \ . \tag{28}$$

Example 11.1. Prove (28) and find the mean and variance of a binomial (n, p) using the moment generating function.

12 Inequalities

12.1 Markov Inequality

Let X be a RV such that $P(X \geq 0) = 1$. Then, for any number $t > 0$,

$$P(X \geq t) \leq \frac{E(X)}{t} \ . \tag{29}$$

12.2 Chebyshev Inequality

Let X be a RV for which $Var(X)$ exists. Then, for any number $t > 0$,

$$P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2} \ . \tag{30}$$