Expectation

Chapter 7

Definitions

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

Properties

- Transformations
- Linearity
- Monotonicity
- Expectation and Independence

Recall: μ_X minimizes $E[(X-c)^2]$ w.r.t. c.

The Prediction Problem

The Problem: Let X and Y be JDRVs with finite means and variances: find a and b to minimize

$$MSE = E[(Y - aX - b)^2] \qquad (*)$$

The Solution: For any a, (*) is minimized by

$$b = E(Y - aX) = \mu_Y - a\mu_X,$$

in which case

$$MSE = E[(\tilde{Y} - a\tilde{X})^2],$$

where

$$\tilde{Y} = Y - \mu_Y,$$
$$\tilde{X} = X - \mu_X.$$

So,

$$\begin{aligned} MSE &= E(\tilde{Y}^2) - 2aE(\tilde{X}\tilde{Y}) + a^2E(\tilde{X}^2) \\ &= \sigma_Y^2 - 2a\sigma_{XY} + a^2\sigma_X^2, \end{aligned}$$

where

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

Here

$$\frac{d}{da} MSE = -2\sigma_{XY} + 2a\sigma_X^2.$$

So, MSE is minimum when

$$a = \frac{\sigma_{XY}}{\sigma_X^2}, \quad \text{and} \quad b = \mu_Y - a\mu_X,$$

if $\sigma_X^2 > 0$. With this a and b,

$$\hat{Y} = aX + b$$

is called the *best linear predictor of* Y.

Note: The minimum MSE is

$$\sigma_Y^2 - 2a\sigma_{XY} + a^2\sigma_X^2 = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}.$$

Example

Galton's Data

In this example

$$\begin{aligned} X &= \text{father's height}, \\ Y &= \text{son's height}, \\ \mu_X &= 68'', \\ \mu_Y &= 69'', \\ \sigma_X &= \sigma_Y = 1'', \\ \sigma_{XY} &= .5. \end{aligned}$$

So,

$$a = \frac{1}{2},$$
$$b = 69 - \frac{1}{2}68 = 35,$$

and

$$\hat{Y} = 35 + \frac{1}{2}X = 69 + \frac{1}{2}(X - 68).$$

Regression Effect:

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

Interpretation: The minimum MSE is

$$\sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2} = \sigma_Y^2 \left[1 - \rho^2\right].$$

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

Special Cases: *a)* $\sigma_{XX} = \sigma_X^2$.

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

Note: Measures of dependence.

Example

Sampling

Suppose that a SRS of $n = 2$ is drawn *w.o.r.* from R red and $N - R$ white tickets; and let

$$p = R/N,$$
$$A_i = \{\text{red on the } i^{th} \text{ draw},$$
$$X_i = \mathbf{1}_{A_i}$$

for $i = 1, 2$. Then

$$E(X_i) = P(A_i) = \frac{R}{N} = p,$$
$$Var(X) = pq,$$
$$E(X_1 X_2) = \frac{R(R-1)}{N(N-1)},$$

and

$$\begin{aligned}\sigma_{12} &= E(X_1X_2) - \mu_1\mu_2 \\ &= \frac{R(R-1)}{N(N-1)} - (\frac{R}{N})^2 \\ &= \frac{NR(R-1) - (N-1)R^2}{N^2(n-1)} \\ &= -\frac{R(N-R)}{N^2(N-1)} \\ &= -\frac{pq}{N-1}.\end{aligned}$$

So,

$$\rho = -\frac{pq/(N-1)}{pq} = -\frac{1}{N-1}.$$

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

$$X' = aX + b,$$
$$Y' = cY + d,$$

then

$$E(X') = aE(X) + b,$$
$$E(Y') = cE(Y) + d,$$

$$\begin{aligned} C(X', Y') &= E[(X' - \mu_{X'})(Y' - \mu_{Y'})] \\ &= E[ac(X - \mu_X)(Y - \mu_Y)] \\ &= acC(X, Y). \end{aligned}$$

and $\cdots$

$$\rho_{X'Y'} = \frac{ac}{|ac|}\rho_{XY} = \pm\rho_{XY}$$

Note: $\sigma^2_{X'} = \sigma_{X',X'} = a^2\sigma_X$.

Sums

Theorem. Let $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ be JDRVs with finite means and variances, and let

$$S = X_1 + \cdots + X_m,$$
$$T = Y_1 + \cdots + Y_n.$$

Then

$$E(S) = E(X_1) + \cdots + E(X_m),$$
$$E(T) = E(Y_1) + \cdots + E(Y_n),$$

and

$$C(S, T) = \sum_{i=1}^{m}\sum_{j=1}^{n} C(X_i, Y_j).$$

Proof. $\cdots$

The Variance of Sum

Corollary

$$D^2(S) = \sum_{i=1}^{m} D^2(X_i) + 2\sum\sum_{i\neq j} C(X_i, X_j).$$

Proof. $D^2(S) = C(S, S)$.

Def: $X_1, \cdots, X_m$ are *uncorrelated* if $C(X_i, X_j) = 0$ for all $i \neq j$.

Corollary. If $X_1, \cdots, X_m$ are uncorrelated, then

$$D^2(S) = D^2(X_1) + \cdots + D^2(X_m). \qquad (*)$$

In particular, (*) holds if $X_1, \cdots, X_m$ are independent.

Example: Binomial.

The Signal Plus Noise Problem

Suppose that

$$X = Y + Z,$$

where

$$Y \perp Z \quad \text{and} \quad E(Z) = 0.$$

Then

$$E(Y) = E(X),$$
$$\sigma^2_X = \sigma^2_Y + \sigma^2_Z,$$
$$\sigma_{XY} = \sigma_{YY} + \sigma_{YZ} = \sigma^2_Y,$$

and

$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\sigma_Y}{\sqrt{\sigma^2_Y + \sigma^2_Z}}.$$

Best Linear Predictor: Find

$$\hat{Y} = (1 - \rho^2)\mu_Y + \rho^2 X.$$

Conditional Expectation

If $X, Y \sim f$ and $f_X(x) > 0$, then

$$E(Y|X=x) = \sum_{y \in \mathcal{Y}} y f_{Y|X}(y|x),$$

in the discrete case, or

$$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

in the continuous case, where where

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)},$$

provided that the sum or integral converges.

Example

In a bridge game, South has five spades. How many spades does North have? Let

$$X = \#\text{spades in South's hand}$$
$$Y = \#\text{spades in North's hand}$$

Then

$$f_{Y|X}(y|X=5) = \frac{\binom{8}{y}\binom{31}{13-y}}{\binom{39}{13}}$$

for $y = 0, \cdots, 8$. So,

$$E(Y|X=5) = 13 \times (\frac{8}{39}).$$

Example

If

$$Y \sim \text{Unif}[0,1] \quad \text{and} \quad X|Y \sim \text{Unif}[0,y]$$

what is $E(Y|X=\frac{1}{2})$? Here

$$f_Y(y) = 1$$

$0 < y \leq 1$ and

$$f_{X|Y}(x|y) = \frac{1}{y}$$

for $0 < x \leq y \leq 1$. So,

$$f(x,y) = f_Y(y) f_{X|Y}(x|y) = \frac{1}{y}$$

for $0 < x \leq y \leq 1$,

$$f_X(x) = \int_x^1 \frac{dy}{y} = \log(y)|_{y=x}^1 = -\log(x),$$

$$f_{Y|X}(y|x) = -\frac{1}{y\log(x)},$$

and

$$\begin{aligned} E(Y|X=\frac{1}{2}) &= \int_{\frac{1}{2}}^1 y f_{Y|X}(y|\frac{1}{2}) dy \\ &= -\int_{\frac{1}{2}}^1 \frac{dy}{\log(\frac{1}{2})} \\ &= -\frac{1}{2\log(\frac{1}{2})} \\ &= \frac{1}{2\log(2)} \end{aligned}$$

Smoothing

Notation: Write $E(Y|X)$ when x is replaced by X.

Theorem. If Y has a finite expectation, then

$$E(Y) = E[E(Y|X)].$$

Proof. See the text.

Example: *The Trapped Miner.* Two doors:

- One leads to safety after 3 hours;
- Two leads back to the mine after 5.

Let Y be the time required to get back. Then $E(Y|X=1) = 3$, and $E(Y|X=2) = 5+3 = 8$. If the miner chooses a door at random, then

$$E(Y) = \frac{1}{2}E(Y|X=1) + \frac{1}{2}E(Y|X=2) = 5.5.$$

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 in some non-degenerate interval.

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

$$\begin{aligned} M(t) &= \int_0^\infty e^{tx}\lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty e^{-(\lambda - t)x} dx \\ &= -\lambda \frac{e^{-(\lambda-t)x}}{\lambda - t}\Big|_{x=0}^{\infty} \\ &= \frac{\lambda}{\lambda - t}, \end{aligned}$$

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

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

Theorem. If $M(t) < \infty$ for $|t| < h$ for some $h > 0$, then

$$\begin{aligned} M(0) &= 1, \\ M'(0) &= \mu, \\ M''(0) &= \mu_2, \\ &\cdots \\ M^{(k)}(0) &= \mu_k, \end{aligned}$$

for all $k = 0, 1, 2, \cdots$.

Proof-Outline-The Discrete Case. If X has PMF f, then

$$M(t) = \sum_{x \in \mathcal{X}} e^{tx} f(x),$$
$$M(0) = \sum_{x \in \mathcal{X}} 1 \times f(x) = 1,$$
$$M'(t) = \sum_{x \in \mathcal{X}} x e^{tx} f(x),$$
$$M'(0) = \sum_{x \in \mathcal{X}} x f(x) = \mu,$$
$$\cdots$$

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 $k = 1, 2, \cdots$.

Corollary

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

Theorem. 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

$$\begin{aligned} M_S(t) &= E(e^{tS}) \\ &= E(\prod_{i=1}^{n} e^{tX_i}) \\ &= \prod_{i=1}^{n} E(e^{tX_i}) \\ &= \prod_{i=1} M_i(t). \end{aligned}$$

Note: Product, not convolution.

Unicity

Theorem: If

$$M_X(t) = M_Y(t)$$

for all t in some non-degerate interval, then

$$F_X(z) = F_Y(z)$$

for all z.

Example. If

$$M(t) = \cosh(t) = \frac{e^t + e^{-t}}{2},$$

then

$$P[X = \pm 1] = \frac{1}{2}.$$

Example

Normal

If $X_1, \cdots, X_n$ are independent and

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

then

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

where

$$\begin{aligned} \mu &= \mu_1 + \cdots + \mu_n, \\ \sigma^2 &= \sigma_1^2 + \cdots + \sigma_n^2, \end{aligned}$$

since

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