Statistical Foundation of Data Science (MA 589)

Lecture Slides

Topic 03: Random Vector

Jointly Distributed Random Variables

Definition 3.1: A function $\boldsymbol{X} : \mathcal{S} \to \mathbb{R}^n$ is called a random vector.

Definition 3.2: For any random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, the joint cumulative distribution function (JCDF) is defined by

$$F_{\boldsymbol{X}}(x_1, \ldots, x_n) = P\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right),$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$.

Remark 3.1: $F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y)$.

Remark 3.2: $F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y)$.

Properties of JCDF

1. $\lim_{x\to\infty}\lim_{y\to\infty} F_{X,Y}(x, y) = 1$.
2. $\lim_{x\to-\infty} F_{X,Y}(x, y) = 0$ for all $y \in \mathbb{R}$.
3. $\lim_{y\to-\infty} F_{X,Y}(x, y) = 0$ for all $x \in \mathbb{R}$.
4. $F_{X,Y}(\cdot, \cdot)$ is right continuous in each argument keeping other fixed.
5. For $-\infty < a_1 < b_1 < \infty$ and $-\infty < a_2 < b_2 < \infty$,

$$F_{X,Y}(b_1, b_2) - F_{X,Y}(b_1, a_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, a_2) \geq 0.$$

Theorem 3.1: Let $G : \mathbb{R}^2 \to \mathbb{R}$ be a function satisfying above properties. Then G is a JCDF of some 2-dimensional random vector.

Discrete Random Vector

Definition 3.3: A random vector (X, Y) is said to have a discrete distribution if there exists an atmost countable set $S_{X,Y} \in \mathbb{R}^2$ such that $P((X, Y) = (x, y)) > 0$ for all $(x, y) \in S_{X,Y}$ and $P((X, Y) \in S_{X,Y}) = 1$. $S_{X,Y}$ is called the support of (X, Y).

Definition 3.4: Define a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ by

$$f_{X,Y}(x, y) = \begin{cases} P(X = x, Y = y) & \text{if } (x, y) \in S_{X,Y} \\ 0 & \text{otherwise.} \end{cases}$$

The function $f_{X,Y}$ is called joint probability mass function (JPMF) of the DRV (X, Y).

Properties of JPMF

1. $f_{X,\,Y}(x,\, y) \geq 0$ for $(x,\, y) \in \mathbb{R}^2$.
2. $\displaystyle\sum_{(x,\,y)\in S_{X,\,Y}} f_{X,\,Y}(x,\, y) = 1.$

Theorem 3.2: If a function $g : \mathbb{R}^2 \to \mathbb{R}$ satisfy 1 and 2 above for the atmost countable set $D = \{(x,\, y) \in \mathbb{R}^2 : g(x,\, y) > 0\}$ in place of $S_{X,\,Y}$, then g is JPMF of some 2-dimensional DRV.

Examples

Find the value of c in the following cases.

Example 3.1: Let (X, Y) be a DRV with JPMF

$$f_{X,Y}(x, y) = \begin{cases} cy & \text{if } x = 1, 2, \ldots, n;\ y = 1, 2, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.2: Let (X, Y) be a DRV with JPMF

$$f_{X,Y}(x, y) = \begin{cases} cy & \text{if } x = 1, 2, \ldots, n;\ y = 1, 2, \ldots, n;\ x \leq y \\ 0 & \text{otherwise.} \end{cases}$$

Marginal PMF from JPMF

Theorem 3.3: Let $(X,\ Y)$ be a discrete random vector with JPMF $f_{X,\,Y}(\cdot,\ \cdot)$ and support $S_{X,\,Y}$. Then X and Y are DRVs. The PMF of X is

$$f_X(x) = \sum_{(x,\,y)\in S_{X,\,Y}} f_{X,\,Y}(x,\ y) \text{ for all fixed } x \in \mathbb{R}. \tag{1}$$

The PMF of Y is given by

$$f_Y(y) = \sum_{(x,\,y)\in S_{X,\,Y}} f_{X,\,Y}(x,\ y) \text{ for all fixed } y \in \mathbb{R}. \tag{2}$$

In this context, $f_X(\cdot)$ and $f_Y(\cdot)$ are called marginal PMF of X and marginal PMF of Y, respectively.

Expectation of Function of DRV

Definition 3.5: Let (X, Y) be a DRV with JPMF $f_{X,Y}$ and support $S_{X,Y}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$. Then the expectation of $h(X, Y)$ is defined by

$$E\left(h(X, Y)\right) = \sum_{(x,y)\in S_{X,Y}} h(x, y) f_{X,Y}(x, y),$$

provided $\displaystyle\sum_{(x,y)\in S_{X,Y}} |h(x, y)| f_{X,Y}(x, y) < \infty.$

Examples

Find the marginal PMFs of X and Y. Also, find $E(X)$, $E(Y)$, and $E(XY)$.

Example 3.3: Let $(X,\ Y)$ be a DRV with JPMF

$$f_{X,Y}(x,\ y) = \begin{cases} cy & \text{if } x = 1,\ 2,\ \ldots,\ n;\ y = 1,\ 2,\ \ldots,\ n \\ 0 & \text{otherwise.} \end{cases}$$

Example 3.4: Let $(X,\ Y)$ be a DRV with JPMF

$$f_{X,Y}(x,\ y) = \begin{cases} cy & \text{if } x = 1,\ 2,\ \ldots,\ n;\ y = 1,\ 2,\ \ldots,\ n;\ x \leq y \\ 0 & \text{otherwise.} \end{cases}$$

Continuous Random Vector

Definition 3.6: A random vector (X, Y) is said to have a continuous distribution if there exists a non-negative integrable function $f_{X,\,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_{X,Y}(x,\, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(t,\, s) ds dt$$

for all $(x,\, y) \in \mathbb{R}^2$.

Definition 3.7: The function $f_{X,\,Y}$ is called the joint probability density function (JPDF) of $(X,\, Y)$.

Definition 3.8: The set $S_{X,\,Y} = \{(x,\, y) \in \mathbb{R}^2 : f_{X,Y}(x,\, y) > 0\}$ is called the support of $(X,\, Y)$.

Properties of JPDF

1. $f_{X,\,Y}(x,\; y) \geq 0$ for $(x,\; y) \in \mathbb{R}^2$.
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,\,Y}(x,\; y)dxdy = 1.$

Theorem 3.4: If a function $g : \mathbb{R}^2 \to \mathbb{R}$ satisfy 1 and 2 above, then g is JPDF of some 2-dimensional CRV.

JPDF to Marginal PDF

Theorem 3.5: Let (X, Y) be a continuous random vector with JPDF $f_{X,Y}(\cdot, \cdot)$. Then X and Y are CRVs. The PDF of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy \quad \text{for all fixed } x \in \mathbb{R}.$$

The PDF of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx \quad \text{for all fixed } y \in \mathbb{R}.$$

In the context of continuous random vector, $f_X(\cdot)$ and $f_Y(\cdot)$ are called marginal PDF of X and marginal PDF of Y, respectively.

Expectation of Function of CRV

Definition 3.9: Let (X, Y) be a CRV with JPDF $f_{X,Y}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$. Then the expectation of $h(X, Y)$ is defined by

$$E(h(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) f_{X,Y}(x, y) dxdy,$$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| f_{X,Y}(x, y) dxdy < \infty$.

Examples

Example 3.5: Let $(X,\ Y)$ be a CRV with JPDF

$$f_{X,Y}(x,\ y) = \begin{cases} ce^{-(2x+3y)} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Examples

Example 3.6: $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$ for real constants a_i.

Example 3.7: At a party N men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who selects their own hat.

Some Remarks

Remark 3.3: (X, Y) is discrete random vector iff X and Y are discrete random variables.

Remark 3.4: If (X, Y) is continuous random vector, then X and Y are continuous random variables.

Remark 3.5: If (X, Y) is continuous random vector, then

$$P\left((X, Y) \in A\right) = \int\int_{(x,y)\in A} f_{X,Y}(x, y)dxdy,$$

for all $A \subseteq \mathbb{R}^2$ such that the integration is possible.

Some Remarks

Remark 3.6: (X, Y) may not be a continuous random vector even if X and Y are continuous random variables.

Remark 3.7: In general, if there exists a set A in $\mathbb{R}^2$ whose area is zero and $P((X, Y) \in A) > 0$, then (X, Y) does not have a JPDF.

Remark 3.8: If the joint distribution is known, then the marginal distributions can be recovered. However, the converse is not true.

Example 3.8: Let f and g be two PDFs and F and G be the corresponding CDFs. Define, for $-1 < \alpha < 1$,

$$h(x, y) = f(x)g(y)\left\{1 + \alpha(1 - 2F(x))(1 - 2G(y))\right\}.$$

Then h is a JPDF whose marginals are f and g.

Independent Random Variables

Definition 3.10: The random variables $X_1, X_2, \ldots, X_n$ are said to be independent if

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

Remark 3.9: X and Y are independent iff the events $E_x = \{X \leq x\}$ and $F_y = \{Y \leq y\}$ are independent for all $(x, y) \in \mathbb{R}^2$.

Remark 3.10: For DRV/CRV (X, Y), the condition of independence is equivalent to

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) \text{ for all } (x, y) \in \mathbb{R}^2.$$

Independent Random Variables

Theorem 3.6: If X and Y are independent, then

$$E\left(g(X)h(Y)\right) = E\left(g(X)\right)E\left(h(Y)\right),$$

provided all the expectations exist.

Covariance and Correlation

Definition 3.11: The covariance of two random variables X and Y is defined by

$$Cov(X, Y) = E\left[(X - E(X))(Y - E(Y))\right] = E(XY) - E(X)E(Y).$$

Definition 3.12: The correlation coefficient of X and Y is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}.$$

Some Properties

1. If X and Y are independent, then $Cov(X, Y) = 0$. The converse is not true in general.
2. $|\rho(X, Y)| \leq 1$.
3. $Cov(X, X) = Var(X)$.
4. $Cov(X, Y) = Cov(Y, X)$.
5. $Cov(aX, Y) = aCov(X, Y)$.
6. $Cov(X + Z, Y) = Cov(X, Y) + Cov(Z, Y)$.
7. $Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)$.
8. $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, Y_j)$.
9. If X_i's are independent, then $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$.

Functions of Random Variables: Technique 1

Example 3.9: Let X_1 and X_2 be *i.i.d.* $U(0, 1)$ random variables. Find the CDF of $Y = X_1 + X_2$.

Example 3.10: Let the JPDF of (X_1, X_2) be given by

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} e^{-x_1} & \text{if } 0 < x_1 < x_2 < \infty \\ 0 & \text{otherwise.} \end{cases}$$

Find the JCDF of $Y_1 = X_1 + X_2$ and $Y_2 = X_2 - X_1$.

Functions of RVs: Technique 2 for DRV

Theorem 3.7: Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a DRV with JPMF $f_{\boldsymbol{X}}$ and support $S_{\boldsymbol{X}}$. Let $g_i : \mathbb{R}^n \to \mathbb{R}$ for all $i = 1, 2, \ldots, k$. Let $Y_i = g_i(\boldsymbol{X})$ for $i = 1, 2, \ldots, k$. Then $\boldsymbol{Y} = (Y_1, , \ldots, Y_k)$ is a DRV with JPMF

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_k) = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in A_{\boldsymbol{y}}} f_{\boldsymbol{X}}(\boldsymbol{x}) & \text{if } (y_1, \ldots, y_k) \in S_{\boldsymbol{Y}} \\ 0 & \text{otherwise,} \end{cases}$$

where $A_{\boldsymbol{y}} = \{\boldsymbol{x} \in S_{\boldsymbol{X}} : g_i(\boldsymbol{x}) = y_i,\ i = 1, \ldots, k\}$ and $S_{\boldsymbol{Y}} = \{(g_1(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x})) : \boldsymbol{x} \in S_{\boldsymbol{X}}\}$.

Functions of RVs: Technique 2 for DRV

Example 3.11: $X_1 \sim P(\lambda_1)$ and $X_2 \sim P(\lambda_2)$ and they are independent. Then $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

Example 3.12: $X_1 \sim Bin(n_1, p)$ and $X_2 \sim Bin(n_2, p)$ and they are independent. Then $X_1 + X_2 \sim Bin(n_1 + n_2, p)$.

Functions of RVs: Technique 2 for CRV

Theorem 3.8: Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a CRV with JPDF $f_{\boldsymbol{X}}$.

1. Let $y_i = g_i(\boldsymbol{x})$, $i = 1, 2, \ldots, n$ be $\mathbb{R}^n \to \mathbb{R}$ functions such that $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$ is one-to-one. That means that there exists the inverse transformation $x_i = h_i(\boldsymbol{y})$, $i = 1, 2, \ldots, n$ defined on the range of the transformation.
2. Assume that both the mapping and its' inverse are continuous.
3. Assume that partial derivatives $\frac{\partial x_i}{\partial y_j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n$, exist and are continuous.
4. Assume that the Jacobian of the inverse transformation
$$J \doteq \det\left(\frac{\partial x_i}{\partial y_j}\right)_{i,j=1,2,\ldots,n} \neq 0$$
on the range of the transformation.

Then $\boldsymbol{Y} = (g_1(\boldsymbol{X}), \ldots, g_n(\boldsymbol{X}))$ is a CRV with JPDF

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(h_1(\boldsymbol{y}), \ldots, h_n(\boldsymbol{y}))|J|.$$

Functions of RVs: Technique 2 for CRV

Example 3.13: Let X_1 and X_2 be *i.i.d.* $U(0, 1)$ random variables. Find the JPDF of $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

Example 3.14: Let X_1 and X_2 be *i.i.d.* $N(0, 1)$ random variables. Find the PDF of $Y_1 = X_1/X_2$.

Remark 3.11: If X and Y are independent, then $g(X)$ and $h(Y)$ are also independent.

Moment Generating Function

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a RV. The moment generating function (MGF) of $\boldsymbol{X}$ at $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ is defined by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left(\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right),$$

provided the expectation exists.

Theorem 3.9: $E(X_1^{r_1} X_2^{r_2} \cdots X_n^{r_n}) = \frac{\partial^{r_1+r_2+\ldots+r_n}}{\partial t_1^{r_1} \partial t_2^{r_2} \ldots \partial t_n^{r_n}} M_{\boldsymbol{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=0}$.

Theorem 3.10: X and Y are independent iff for all (t_1, t_2) in a neighborhood of origin

$$M_{X,Y}(t_1, t_2) = M_X(t_1) M_Y(t_2).$$

Technique 3

Definition 3.13: Two RVs $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to have the same distribution, denoted by $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$, if $F_{\boldsymbol{X}}(\cdot) = F_{\boldsymbol{Y}}(\cdot)$.

Theorem 3.11: Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two RVs. Let $M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{\boldsymbol{Y}}(\boldsymbol{t})$ for all $\boldsymbol{t}$ in a neighborhood around 0, then $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}$.

Example 3.15: Let X_i, $i = 1, 2, \ldots, k$ be independent $Bin(n_i, p)$ RVs. Then $\sum X_i \sim Bin(\sum n_i, p)$.

Example 3.16: Let X_i, $i = 1, 2, \ldots, k$ be iid $Exp(\lambda)$ RVs. Then $\sum X_i \sim Gamma(k, \lambda)$.

Example 3.17: Let X_i, $i = 1, 2, \ldots, k$ be independent $N(\mu_i, \sigma_i^2)$ RVs. Then $\sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$.

Expectation and Variance of a Random Vector

Expectation of a random vector is given by

$$E(\boldsymbol{X}) = (EX_1, EX_2, \ldots, EX_n) = \boldsymbol{\mu}\,.$$

The variance-covariance matrix of a n-dimensional random vector, denoted by Σ, is defined by

$$\Sigma = [Cov(X_i, X_j)]_{i,j=1}^{n} = E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^t\,.$$

Conditional Distribution for DRV

Definition 3.14: Let (X, Y) be a DRV with JPMF $f_{X,Y}(\cdot,\cdot)$. Suppose the marginal PMF of Y is $f_Y(\cdot)$. The conditional PMF of X, given $Y = y$ is defined by

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

provided $f_Y(y) > 0$.

Remark 3.12:

1. Note that $f_{X,Y}(x, y) = P(X = x, Y = y)$ and $f_Y(y) = P(Y = y)$.
2. Thus, the conditional PMF of X given $Y = y$ is $P(X = x|Y = y)$.

Example 3.18: Let $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$ and X and Y are independent. Find the conditional PMF X given $X + Y = n$.

Conditional CDF for DRV

Definition 3.15: The conditional CDF of X given $Y = y$ is defined by

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{\{u \leq x : (u,y) \in S_{X,Y}\}} f_{X|Y}(u|y) \, .$$

provided $f_Y(y) > 0$.

Theorem 3.12:

1. If X and Y are independent DRVs, then $f_{X|Y}(x|y) = f_X(x)$ for all $x \in \mathbb{R}$ and $y \in S_Y$.
2. If X and Y are independent DRVs, then $F_{X|Y}(x|y) = F_X(x)$ for all $x \in \mathbb{R}$ and $y \in S_Y$.

Conditional Expectation for DRV

Definition 3.16: The conditional expectation of $h(X)$ given $Y = y$ is defined by

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

provided it is absolutely summable.

Remark 3.13:

1. Notice that conditional PMF is a PMF.
2. The conditional expectation is an expectation with respect to conditional PMF.
3. Thus, conditional expectation satisfies all the properties of expectation.

Examples

Example 3.19: Let $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$ and X and Y are independent. Calculate the conditional expectation of X given $X + Y = n$.

Example 3.20: Suppose a system has n components. Suppose on a rainy day, component i functions with probability p_i, $i = 1, 2, \ldots, n$ independent of others. Calculate the conditional expected number of components that will function tomorrow given that it will rain tomorrow.

Conditional Distribution for CRV

Let (X, Y) be a CRV. The conditional CDF of X given $Y = y$ is defined as

$$F_{X|Y}(x|y) = \lim_{\epsilon \downarrow 0} P(X \leq x | Y \in (y - \epsilon, y + \epsilon]) .$$

provided the limit exists.

Define the conditional PDF of X given $Y = y$, $f_{X|Y}(x|y)$, as the non-negative function satisfying

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(t|y) dt, \qquad \forall x \in \mathbb{R} .$$

Conditional Distribution for CRV (Contd.)

Theorem 3.13: Let $f_{X,Y}$ be the JPDF of (X, Y) and let f_Y be the marginal PDF of Y. If $f_Y(y) > 0$, then the conditional PDF of X given $Y = y$ exists and is given by

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

Definition 3.17: The conditional expectation of $h(X)$ given $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

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

provided it is absolutely integrable.

Examples

Example 3.21: Suppose the JPDF of (X, Y) is given by

$$f_{X,Y}(x,y) = \begin{cases} 6xy(2-x-y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.

Example 3.22: $f_{X,Y}(x,y) = \frac{1}{2}ye^{-xy}$, $0 < x < \infty, 0 < y < 2$. Find $E(e^{X/2}|Y = 1)$.

Properties of Conditional Expectation

Suppose either (X, Y) is a DRV or a CRV. Define $E(X|Y) = g(Y)$, where $g(y) = E(X|Y = y)$. Thus $E(X|Y)$ is again a random variable.

Theorem 3.14: $E(X) = E(E(X|Y))$.

Theorem 3.15: $E(X - E(X|Y))^2 \leq E(X - f(Y))^2$ for any function f. Thus $E(X|Y)$ is the "best estimate of X given Y".

Example 3.23: Virat will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability and history book is Poisson with mean 2 and 5 respectively, then assuming that Virat is equally likely to choose either book, what is the expected number of misprints that he will come across?

Conditional Variance

Definition 3.18: Let (X, Y) be a random vector.
$Var(X|Y) = h(Y)$ where $h(y) = E((X - E(X|Y))^2|Y = y)$
$= E(X^2|Y = y) - (E(X|Y = y))^2$.

Theorem 3.16: $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$.

Example 3.24: Let $X_0, X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with mean μ and variance σ^2. Let $N \sim Bin(n, p)$, independent of $\{X_i\}$.

Define $S = \sum_{i=0}^{N} X_i$. Find $Var(S)$.

Computing Probability by Conditioning

$$P(E) = \begin{cases} \sum_y P(E|Y=y)P(Y=y) & \text{for} \quad Y \text{ discrete} \\ \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)dy & \text{for} \quad Y \text{ continuous}\,. \end{cases}$$

Example 3.25: Let X and Y be independent CRVs having PDFs f_X and f_Y, respectively. Compute $P(X < Y)$.

Example 3.26: Let X and Y be i.i.d. CRVs having common PDF f. Then $P(X < Y) = P(X > Y) = 0.5$. And $P(X = Y) = 0$.

Example 3.27: Suppose X and Y are two independent RVs, either discrete or continuous. What is the distribution of $X + Y$?

Conditional Expectation for given Event

Definition 3.19: Let (X, Y) be a random vector. Then

$$E\left(h(X,\ Y) | (X,\ Y) \in A\right) = \frac{E(h(X,\ Y) I_A(X,\ Y))}{P\left((X,\ Y) \in A\right)}.$$

Example 3.28: $X \sim Exp(1)$. Find $E\left(X | X \geq 2\right)$.

Example 3.29: $(X,\ Y)$ is uniform on unit square. Find $E\left(X | X + Y > 1\right)$.

Example 3.30: A rod of length l is broken into two parts. Find the expected length of the shorter part.