Study Guide for Joint Distributions (PSTAT 120A/B)

Botao Jin

University of California, Santa Barbara — May 30, 2024

Definition of Joint Distributions

1. If X, Y are two random variables, then he random vector $(X, Y) : \Omega \to \mathbb{R}^2$.
2. If we looked at n random variables $X_1, \dots, X_n$ jointly, then random vectors $(X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$.
3. The probability distribution of $(X_1, \dots, X_n)$ is represented by $P((X_1, \dots, X_n) \in B)$ for B is a subset of $\mathbb{R}^n$.
4. In this study guide, we only only consider the joint distribution of two dimensional cases (X, Y). You can check lecture notes of 120A for more detail about n dimensional cases, but they are quite similar.

Discrete Joint Distributions

1. Definition of joint PMF: $P_{X,Y}(x, y) = P(X = x, Y = y) = P(\{X = x\} \cap \{Y = y\})$.
2. $P_{X,Y}(x, y) \geq 0$ for all possible x, y.
3. $\sum_x \sum_y P_{X,Y}(x, y) = 1$.
4. Expected values $\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) P_{X,Y}(x, y)$.
5. Special Case (Multinomial distribution):
   - Joint distribution for $(X_1, X_2, \dots, X_r)$.
   - Parameters n, r and $p_1, \dots, p_r$ with $\sum_{i=1}^r p_i = 1$.
   - Support: integer vectors $(k_1, \dots, k_r)$ in which $k_i \geq 0$ and $k_1 + \dots + k_r = n$
   - Joint PMF:
   $$P(X_1 = k_1, \dots, X_r = k_r) = \binom{n}{k_1, \dots, k_r} p_1^{k_1} \dots p_r^{k_r}$$
   - Special case: Binomial distribution (Cases for $r = 2$).

Continuous Joint Distributions

1. Joint PDF: $f_{X,Y}(x, y)$ satisfies:
   - $\iint_{\mathbb{R}^2} f_{X,Y}(x, y) dxdy = 1$
   - $f_{X,Y}(x, y) \geq 0$
   - $\iint_B f_{X,Y}(x, y) dxdy = P((X, Y) \in B)$
2. $\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy = \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dydx$.
3. Special case (Uniform distribution in two dimensions): $(X, Y) \sim$ Uniform(D), for $D \subseteq \mathbb{R}^2$ with $|D| =$ Area$(D) < \infty$, then the density function for (X, Y) is
$$f(x, y) = \begin{cases} 1/|D| & (x, y) \in D \\ 0 & (x, y) \notin D \end{cases}$$

Marginal Distributions

1. Discrete case:
   - Distribution of X: $p_X(x) = P(X = x) = \sum_y P_{X,Y}(x, y)$
   - Distribution of Y: $p_Y(y) = P(Y = y) = \sum_x P_{X,Y}(x, y)$
2. Continuous Case:
   - Distribution of X: $f(x) = \int_{\mathbb{R}} f(x, y)dy$
   - Distribution of Y: $f(y) = \int_{\mathbb{R}} f(x, y)dx$
3. For general X, Y, we have joint CDF $F(x, y) = P(X \leq x, Y \leq y)$ with
   - Marginal CDF of X: $F(x) = \lim_{y\to\infty} F(x, y)$
   - Marginal CDF of Y: $F(y) = \lim_{x\to\infty} F(x, y)$

Joint Distributions for independent RVs

1. Two dimensional case: Suppose (X, Y) be a random vector with which X and Y are independent
   - If (X, Y) is discrete, then the joint PMF is $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for p_X and p_Y be marginal PMF of X and Y, respectively.
   - If (X, Y) is continuous, then the joint PDF is $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for f_X and f_Y be marginal PDF of X and Y, respectively.
   - For general X, Y: for any subsets B_1, B_2 of $\mathbb{R}$, we have $P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$. In particular, joint CDF $F(x, y) = F_X(x)F_Y(y)$ with marginal CDF F_X and F_Y of X and Y, respectively.
2. Suppose n dimension vector $(X_1, \ldots, X_n)$ with which $X_1, \ldots, X_n$ are independent:
   - Discrete: $P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{k=1}^{n} P(X_k = x_k)$
   - Continuous: joint PDF $f(x_1, \ldots, x_n) = \prod_{k=1}^{n} f_{X_k}(x_k)$, i.e. the product of marginal PDFs of X_k
   - General case: Joint CDF $F(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_{X_k}(x_k)$, i.e. the product of marginal CDFs of X_i.
3. Sum of n random variables:

   (a) For some special distributions:
   - Suppose $X_1 \sim Poisson(\lambda_1), \ldots, X_n \sim Poisson(\lambda_n)$, and $X_1, \ldots, X_n$ are **independent**, then
     $$X_1 + \cdots + X_n \sim Poisson(\lambda_1 + \cdots + \lambda_n)$$
     which implies that the sum of independent Poisson random variables still follows Poisson distribution.
   - For independent Bernoulli random variables $X_1, \ldots, X_n \sim Ber(p)$, we have
     $$X_1 + \cdots + X_n \sim Bin(n, p)$$
     In particular, if $X \sim Bin(n_1, p)$, $Y \sim Bin(n_2, p)$, and X and Y are independent, then $X + Y \sim Bin(n_1 + n_2, p)$.
   - For independent normal random variables $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, n$, where $a_i, b \in \mathbb{R}$, then
     $$X = a_1X_1 + \cdots + a_nX_n + b \sim N(\mu, \sigma^2)$$
     with $\mu = a_1\mu_1 + \cdots + a_n\mu_n + b$ and $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$.

(b) Linearity of Expectation: For random variables $X_1, \ldots, X_n$ and real numbers $a_1, \ldots, a_n$:

$$E[a_1X_1 + \cdots + a_nX_n] = a_1E[X_1] + \cdots + a_nE[X_n]$$

Remark. In this case, the independence of $X_1, \ldots, X_n$ is **NOT** needed.

(c) Product of expectation: For independent random variables $X_1, \ldots, X_n$, we have

$$E\left[\prod_{k=1}^{n} X_n\right] = \prod_{k=1}^{n} E[X_k]$$

In particular, the moment generating function in this case is

$$M_{X_1+\cdots+X_n}(t) = E[\exp\{t(X_1 + \cdots + X_n)\}] = E\left[\prod_{k=1}^{n} e^{tX_k}\right] = \prod_{k=1}^{n} E[e^{tX_k}] = \prod_{k=1}^{n} M_{X_k}(t)$$

(d) Sum of variance: For independent random variables $X_1, \ldots, X_n$, we have

$$Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$$

Covariance

1. Let X, Y be two random variables with mean μ_1, μ_2, then the covariance of X and Y is

$$Cov(X, Y) = E[(X - \mu_1)(Y - \mu_2)] = E[XY] - E[X]E[Y]$$

2. If X, Y are independent, then $Cov(X, Y) = 0$. But the converse is not true. For example, X has a PMF $P(X = 1) = P(X = 0) = 1/2$, consider X and X^3.

   Remark. If both X and Y are Gaussian (normal) distribution, then $Cov(X, Y) = 0$ implies X and Y are independent.

3. Variance of sum of random variables $X_1, \ldots, X_n$:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{1\le i<j\le n} Cov(X_i, X_j)$$

4. Correlation: For random variables X and Y,

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

.

Conditional distribution

1. Let X be a discrete random variables and B be an event with $P(B) > 0$, then conditional PMF of X on B is

$$p_{X|B}(k) = P(X = k|B) = \frac{P(\{X = k\} \cap B)}{P(B)}$$

   for $k \in S_X$ with S_X be a support of X.

   - Conditional expectation:

$$E[X|B] = \sum_{k\in S_X} kp_{X|B}(k) = \sum_{k\in S_X} kP(X = k|B)$$

- Suppose $B_1, \dots, B_n$ be a partition of sample space Ω, then by Law of total probability:

$$P(X = k) = \sum_{i=1}^{n} P(X = k|B_i)P(B_i)$$

for $k \in S_X$ and the expected value

$$\begin{aligned} E[X] &= \sum_{k \in S_x} kP(X = k) \\ &= \sum_{k \in S_x} k \sum_{i=1}^{n} P(X = k|B_i)P(B_i) \\ &= \sum_{i=1}^{n} P(B_i) \sum_{k \in S_x} kP(X = k|B_i) \\ &= \sum_{i=1}^{n} E[X|B_i]P(B_i) \end{aligned}$$

2. Let X, Y are discrete random variables with supports S_X, S_Y and $x \in S_X, y \in S_Y$, the conditional PMF is

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

with conditional expected value: $E[X|Y = y] = \sum_{x \in S_X} x p_{X|Y}(x|y)$ for $y \in S_Y$. Since events $\{Y = y\}$ for $y \in S_Y$ form a partition of Ω, we have

$$E[X] = \sum_{y \in S_Y} E[X|Y = y]P(Y = y)$$

3. Let X, Y are continuous random variables, the conditional density of X given $Y = y$ is

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

with $f_{X,Y}$ joint PDF, f_Y marginal PDF of Y, and conditional expected value: $E[X|Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x|y)dx$

Remark 1 In this case, we have

$$\begin{aligned} E[X] &= \iint x f_{X,Y}(x, y)dxdy \\ &= \iint x f_{X|Y}(x|y) f_Y(y)dxdy \\ &= \iint x f_{X|Y}(x|y)dx f_Y(y)dy \\ &= \int E[X|Y = y] f_Y(y)dy \end{aligned}$$

Remark 2 For general X, Y, $E[X|Y = y]$ is a function of y, rather than X, so $E[X|Y]$ should be a function of Y.

Remark 3 Tower property: $E[E[X|Y]] = E[X]$.

Applications

Suppose that $X_1, X_2, \dots$ be a random sample (i.i.d) with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2$, both μ and σ are finite. Define $\bar{X}_n = (\sum_{i=1}^{n} X_i)/n$, which is the average/mean of first n observations.

1. Weak Law of large numbers: for any $\epsilon > 0$, $P(|\bar{X}_n - \mu| < \epsilon) = 1$ as $n \to \infty$.

2. Strong Law of large number (Optional): $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$.
3. Central Limit Theorem (120B): mean $E[\bar{X}_n] = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$

$$P\left(a < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \to \Phi(b) - \Phi(a)$$

for Φ be CDF of standard normal distribution

Note: This study guide is used for Botao Jin's sections only. Comments, bug reports: b_jin@ucsb.edu