MA2202: Probability I

Random vectors

Spring 2021

Satvik Saha
19MS154

Indian Institute of Science Education and Research, Kolkata,
Mohanpur, West Bengal, 741246, India.

Definition 4.1 (Random vector)**.** A random vector $\boldsymbol{X} \colon \Omega \to \mathbb{R}^n$ is a tuple of random variables $X_i \colon \Omega \to \mathbb{R}$.

Definition 4.2 (Joint cumulative distribution function)**.** The joint cumulative distribution function of a random vector $\boldsymbol{X}$ is the map $F_{\boldsymbol{X}} \colon \mathbb{R}^n \to [0, 1]$, given as

$$F_{\boldsymbol{X}}(\boldsymbol{s}) = P(X_1 \leq s_1, \dots, X_n \leq s_n).$$

Definition 4.3 (Joint probability mass function)**.** If X_i are discrete random variables, their joint probability mass function is the map $p_{\boldsymbol{X}} \colon \mathbb{R}^n \to [0, 1]$,

$$p_{\boldsymbol{X}}(\boldsymbol{s}) = P(X_1 = s_1, \dots, X_n = s_n).$$

Definition 4.4 (Joint probability density function)**.** Suppose that

$$F_{\boldsymbol{X}}(\boldsymbol{s}) = \int_{-\infty}^{s_n} \cdots \int_{-\infty}^{s_1} f_{\boldsymbol{X}}(t_1, \dots, t_n)\, dt_1 \dots\, dt_n,$$

then $f_{\boldsymbol{X}} \colon \mathbb{R}^n \to [0, 1]$ is the probability density function corresponding to the joint cumulative distribution function $F_{\boldsymbol{X}}$.

Remark. If $f_{\boldsymbol{X}}$ is continuous, then

$$f_{\boldsymbol{X}} = \frac{\partial F_{\boldsymbol{X}}(t_1, \dots, t_n)}{\partial t_1 \dots \partial t_n}.$$

Definition 4.5 (Joint moment generating function)**.** Let $\boldsymbol{X}$ be a random vector. Then, its joint moment generating function is defined as

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = E\left[e^{\boldsymbol{t}^\top \boldsymbol{X}}\right] = E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right].$$

Remark. If $X_1, \ldots, X_n$ are independent, then

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \prod M_{X_i}(t_i).$$

Theorem 4.1. *If X and Y are independent continuous random variables, then the probability density function of their sum is the convolution $f_{X+Y} = f_X * f_Y$,*

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_X(x-t)\, f_Y(t)\, dt.$$

Example. When X and Y are identical and uniform on $[0, 1]$, then

$$f_{X+Y}(x) = \int_0^1 f(x-t)\, dt = \begin{cases} x, & \text{if } a \in [0,1], \\ 2-x, & \text{if } a \in [1,2], \\ 0, & \text{otherwise }. \end{cases}$$

Also,

$$M_{X+Y}(t) = (M(t))^2 = \frac{1}{t^2}(e^t - 1)^2.$$

Definition 4.6 (Conditional distribution)**.** Let X and Y be two discrete random variables. We write

$$P(X = s \,|\, Y = t) = \frac{P(X = s, Y = t)}{P(Y = t)}$$

for $P(Y = t) > 0$. We also have

$$P(X \leq s \,|\, Y = t) = \sum_{r \leq s} P(X = r \,|\, Y = t).$$

If X and Y are continuous random variables, then the conditional distribution of X given $Y = t$ is described as

$$F_{X \,|\, Y=t}(r) = \int_{-\infty}^{s} \frac{f_{X,Y}(r,t)}{f_Y(t)}\, dr.$$

Example. Consider two continuous random variables X and Y which have a joint probability mass function

$$f_{X,Y}(s,t) = \begin{cases} \alpha t, & \text{if } 0 < t < s < 1, \\ 0, & \text{otherwise.} \end{cases}$$

First normalize, by demanding

$$\iint_{\mathbb{R}^2} f_{X,Y}(s,t)\, dt\, ds = \int_0^1 \int_0^s \alpha t\, dt\, ds = 1,$$

whence $\alpha = 6$. Thus,

$$E\left[Y \mid X = s\right] = \int_{\mathbb{R}} t \cdot \frac{f_{X,Y}(s,t)}{f_X(s)}\, dt.$$

Now,

$$f_X(s) = \int_{\mathbb{R}} f_{X,Y}(s,t)\, dt = \int_0^s 6t\alpha\, dt = 3s^2$$

for $0 < s < 1$, and simply 3 for $s \geq 1$. Thus,

$$E\left[Y \mid X = s\right] = \int_0^s t \cdot \frac{6t}{3s^2}\, dt = \frac{2}{3}s,$$

in the region $0 < s < 1$. For $s \geq 1$, the expectation becomes $2/3$. Also,

$$\operatorname{Var}[Y \mid X = s] = E\left[Y^2 \mid X = s\right] - E\left[Y \mid X = s\right]^2.$$

The first term is

$$E\left[Y^2 \mid X = s\right] = \int_0^s t^2 \cdot \frac{6t}{3s^2}\, dt = \frac{1}{2}s^2.$$

Thus,

$$\operatorname{Var}[Y \mid X = s] = \frac{1}{2}s^2 - \frac{4}{9}s^2 = \frac{1}{18}s^2.$$

Note that

$$f_Y(t) = \int_{\mathbb{R}} f_{X,Y}(s,t)\, ds = \int_t^1 6t\alpha\, ds = 6t(1-t)$$

in the region $0 < t < 1$. Thus,

$$F_Y(t) = \int_0^t 6t'(1-t')\, dt' = t^2(3-2t)$$

for $0 < t < 1$. $F_Y(t) = 1$ for $t \geq 1$.

Theorem 4.2. *For discrete or continuous random variables X and Y,*

$$E\left[E\left[X \mid Y\right]\right] = E\left[X\right].$$

Proof.

$$E\left[E\left[X \mid Y\right]\right] = \sum_n E\left[X \mid Y = n\right]\, P(Y = n) = \sum_{nm} mP(X = m, Y = n).$$

Reordering the summations, we get

$$\sum_m m \sum_n P(X = m, Y = n) = \sum_m mP(X = m) = E\left[X\right].$$

The proof for discrete random variables is analogous, switching the sums for integrals. ☐

Theorem 4.3. *For random variables X and Y,*

$$\mathrm{Var}[X] = \mathrm{Var}[E\left[X \mid Y\right]] + E\left[\mathrm{Var}[X \mid Y]\right].$$

Proof. Using the previous theorem,

$$\mathrm{Var}[E\left[X \mid Y\right]] = E\left[E\left[X \mid Y\right]^2\right] - E\left[E\left[X \mid Y\right]\right]^2 = E\left[E\left[X \mid Y\right]^2\right] - E\left[X\right]^2,$$

and

$$E\left[\mathrm{Var}[X \mid Y]\right] = E\left[E\left[X^2 \mid Y\right] - E\left[X \mid Y\right]^2\right] = E\left[X^2\right] - E\left[E\left[X \mid Y\right]^2\right].$$

Adding the above gives the desired result. ☐

Definition 4.7 (Order statistics)**.** Let $X_1, \ldots, X_n$ be discrete independent identically distributed random variables, with a common probability mass function. We define

$$X_{(1)} = \min(X_1, \ldots, X_n), \qquad \ldots \qquad X_{(n)} = \max(X_1, \ldots, X_n).$$

Note that we must have

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n-1)} \le X_{(n)}.$$

Lemma 4.4. *If $X_1, \ldots, X_n$ be discrete independent identically distributed random variables, with a common probability mass function, for any permutation σ of $\{1, \ldots, n\}$,*

$$P(X_1 = s_1, \ldots, X_n = s_n) = P(X_1 = s_{\sigma(1)}, \ldots, X_n = s_{\sigma(n)}).$$

Proof. The expressions are both equal to $p(s_1) \ldots p(s_n)$, where p is the common probability mass function. ☐

Theorem 4.5. *Let $X_1, \ldots, X_n$ be discrete independent identically distributed random variables, and let g denote the joint probability mass function of the order statistics.*

$$g(s_1, \ldots, s_n) = \begin{cases} P(X_{(1)} = s_1, \ldots, X_{(n)} = s_n), & \text{if } s_1 \le \cdots \le s_n, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, let $G_{\tilde{s}}$ denote the group of all permutations of $\{s_1, \ldots, s_n\}$. Recall that $|G_{\tilde{s}}| = n!/(r_1! \ldots r_m!)$ where r_i of the s_j's are equal to some t_i. Thus for increasing $s_1, \ldots, s_n$,

$$g(s_1, \ldots, s_n) = \sum_{\sigma \in G_{\tilde{s}}} P(X_1 = \sigma(s_1), \ldots, X_n = \sigma(s_n)) = |G_{\tilde{s}}| P(X_1 = s_1, \ldots, X_n = s_n).$$

This can also be written as

$$g(s_1, \ldots, s_n) = \binom{n}{r_1 \ldots r_m} p(t_1)^{r_1} \ldots p(t_m)^{r_m}.$$

Theorem 4.6. *Let $X_1, \dots, X_n$ be discrete independent identically distributed random variables, and let F denote their common cumulative distribution function. Then,*

$$P(X_{(n)} \leq s) = P(X_1 \leq s, \dots, X_n \leq s) = F(s)^n.$$

Now,

$$P(X_{(n)} = s) = P(X_{(n)} \leq s) - P(X_{(n)} \leq s-1) = F(s)^n - F(s-1)^n.$$

Similarly,

$$P(X_{(1)} \leq s) = 1 - P(X_1 \geq s, \dots, X_n \geq s) = 1 - (1-F(s))^n.$$

Thus,

$$P(X_{(1)} = s) = (1-F(s-1))^n - (1-F(s))^n.$$

Theorem 4.7. *Let $X_1, \dots, X_n$ be continuous independent identically distributed random variables, let f denote their common probability density function, and let g denote their joint probability density function. As before, for any permutation of $\{s_1, \dots, s_n\}$,*

$$g(s_{\sigma(1)}, \dots, s_{\sigma(n)}) = f(s_1) \dots f(s_n).$$

For small $\epsilon > 0$, we can write

$$P\left(s_{\sigma(1)} - \frac{\epsilon}{2} \leq X_1 \leq s_{\sigma(1)} + \frac{\epsilon}{2}, \dots, s_{\sigma(n)} - \frac{\epsilon}{2} \leq X_n \leq s_{\sigma(n)} + \frac{\epsilon}{2}\right) \approx \epsilon^n f(s_1) \dots f(s_n).$$

Therefore, for $s_1 < s_2 < \dots < s_n$, we have

$$P\left(s_{\sigma(1)} - \frac{\epsilon}{2} \leq X_{(1)} \leq s_{\sigma(1)} + \frac{\epsilon}{2}, \dots, s_{\sigma(n)} - \frac{\epsilon}{2} \leq X_{(n)} \leq s_{\sigma(n)} + \frac{\epsilon}{2}\right) \approx n!\epsilon^n f(s_1) \dots f(s_n).$$

Therefore, dividing by ϵ^n and letting $\epsilon \to 0$, we have

$$g(s_1, \dots, s_n) = n! f(s_1)(s_n).$$

Thus, assuming the continuity of f, we have

$$g(s_1, \dots, s_n) = \begin{cases} n! \lim_{\substack{(r_1, \dots, r_n) \to (s_1, \dots, s_n) \\ r_1 < \dots < r_n}} f(r_1) \dots f(r_n), & \textit{if } s_1 \leq \dots \leq s_n, \\ 0 & \textit{otherwise.} \end{cases}$$

Theorem 4.8. *Let $X_1, \dots, X_n$ be continuous independent identically distributed random variables, and let F denote their common cumulative distribution function. Then, like before,*

$$P(X_{(1)} \leq s) = 1 - (1-F(s))^n, \qquad P(X_{(n)} \leq s) = F(s)^n.$$

Thus, the probability density functions are given by

$$f_{X_{(1)}}(s) = \frac{d}{ds} P(X_{(1)} \leq t) = n(1-F(s))^{n-1} f(s),$$
$$f_{X_{(n)}}(s) = \frac{d}{ds} P(X_{(n)} \leq t) = nF(s)^{n-1} f(s).$$