Lecture-06: Transformation of random vectors

1 Functions of random variables

Definition 1.1. Borel measurable sets on a space $\mathbb{R}^n$ is denoted by $\mathcal{B}(\mathbb{R}^n)$ and generated by the collection $(\pi_i^{-1}(-\infty, x] : x \in \mathbb{R}, i \in [n])$. A function $g : \mathbb{R}^n \to \mathbb{R}^m$ is called **Borel measurable** function, if $g^{-1}(B_m) \in \mathcal{B}(\mathbb{R}^n)$ for any $B_m \in \mathcal{B}(\mathbb{R}^m)$.

Proposition 1.2. *Consider a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Suppose $g : \mathbb{R} \to \mathbb{R}$ is function such that $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$, then $g(X)$ is a random variable.*

Proof. We represent $g(X)$ by a map $Y : \Omega \to \mathbb{R}$ such that $Y(\omega) \triangleq (g \circ X)(\omega)$ for all outcomes $\omega \in \Omega$. We further check that for any half open set $B_x = (-\infty, x]$, we have $Y^{-1}(B_x) = (X^{-1} \circ g^{-1})(B_x)$. Since $g^{-1}(B_x) \in \mathcal{B}(\mathbb{R})$, it follows that $Y^{-1}(B_x) \in \mathcal{F}$ by the definition of random variables. □

Example 1.3 (Monotone function of random variables). Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonically increasing function, then $g^{-1}(-\infty, x] \in \mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$. Consider a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, then $Y \triangleq g(X)$ is a random variable with distribution function

$$F_Y(y) = P\{g(X) \leqslant y\} = P\left\{X \leqslant g^{-1}(y)\right\} = F_X(g^{-1}(y)).$$

Here, $g^{-1}(y)$ is the functional inverse, and not inverse image as we have been seeing typically. We can think $g^{-1}(y) = g^{-1}\{y\}$, though this inverse image has at most a single element since g is monotonically increasing.

Example 1.4. Consider a positive random variable $X : \Omega \to \mathbb{R}_+$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be such that $g(x) = e^{-\theta x}$ for all $x \in \mathbb{R}_+$ and some $\theta > 0$. Then, g is monotonically decreasing in X and $x = g^{-1}(y) = -\frac{1}{\theta}\ln y$. This implies that $g^{-1}(-\infty, y] = [-\frac{1}{\theta}\ln y, \infty) \in \mathcal{B}(\mathbb{R}_+)$ for all $y \in \mathbb{R}_+$. Thus g is a measurable function, and $Y = g(X)$ is a random variable.

Proposition 1.5 (Independence of function of random variables). *Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be functions such that $g^{-1}(-\infty, x]$ and $h^{-1}(-\infty, x]$ are Borel sets for all $x \in \mathbb{R}$. Consider independent random variables X and Y defined on the probability space $(\Omega, \mathcal{F}, P)$, then $g(X)$ and $h(Y)$ are independent random variables.*

Proof. For any $u, v \in \mathbb{R}$, we can define inverse images $A_g(u) \triangleq g^{-1}(-\infty, u]$ and $A_h(v) \triangleq h^{-1}(\infty, v]$. Since g, h are Borel measurable, we have $A_g(u), A_h(v) \in \mathcal{B}(\mathbb{R})$. We can write the following outcome set equality for the joint event

$$\{g(X) \leqslant u\} \cap \{h(Y) \leqslant v\} = \left\{X \in g^{-1}(-\infty, u]\right\} \cap \left\{Y \in h^{-1}(-\infty, v]\right\} = X^{-1}(A_g(u)) \cap Y^{-1}(A_h(v)) \in \mathcal{F}.$$

Since X and Y are independent random variables, it follows that $X^{-1}(A_g(u))$ and $Y^{-1}(A_h(v))$ are independent events, and the result follows. □

2 Function of random vectors

Proposition 2.1. *Consider a random vector $X : \Omega \to \mathbb{R}^n$ defined on the probability space $(\Omega, \mathcal{F}, P)$, and a Borel measurable function $g : \mathbb{R}^n \to \mathbb{R}^m$ such that $A_g(y) \triangleq \cap_{j=1}^m \{x \in \mathbb{R}^n : g_j(x) \leqslant y_j\} \in \mathcal{B}(\mathbb{R}^n)$ for all $y \in \mathbb{R}^m$. Then, $g(X) : \Omega \to \mathbb{R}^m$ is a random vector. The joint distribution function $F_Y : \mathbb{R}^m \to [0,1]$ for the vector $Y \triangleq g(X)$ is given by*

$$F_Y(y) = P(X^{-1}(A_g(y))), \quad \text{for all } y \in \mathbb{R}^m.$$

Example 2.2 (Sum of random variables). For a random vector $X : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Define an addition function $+ : \mathbb{R}^n \to \mathbb{R}$ such that $+(x) = \sum_{i=1}^n x_i$ for any $x \in \mathbb{R}^n$. We can verify that $+$ is a Borel measurable function and hence $Y = +(X) = \sum_{i=1}^n X_i$ is a random variable. When $n = 2$ and X is a continuous random vector with density $f_X : \mathbb{R}^2 \to \mathbb{R}_+$, we can write

$$F_Y(y) = P(\{Y \leqslant y\}) = P(\{X_1 + X_2 \leqslant y\}) = \int_{x_1 \in \mathbb{R}} \int_{x_2 \leqslant y - x_1} f_X(x_1, x_2) dx_1 dx_2.$$

By applying a change of variable $(x_1, t) = (x_1, x_1 + x_2)$ and changing the order of integration, we see that

$$F_Y(y) = \int_{t \leqslant y} dt \int_{x_1 \in \mathbb{R}} dx_1 f_X(x_1, t - x_1).$$

When Y is a continuous random vector, we can write

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \int_{x \in \mathbb{R}} f_X(x, y - x) dx.$$

When $X : \Omega \to \mathbb{R}^2$ is an independent vector, then $f_X(x) = f_{X_1}(x_1) f_X(x_2)$ for all $x \in \mathbb{R}^2$. Therefore, the density of the sum $X_1 + X_2$ is given by

$$f_Y(y) = \int_{x \in \mathbb{R}} dx f_{X_1}(x) f_{X_2}(y - x) = (f_{X_1} * f_{X_2})(y),$$

where $* : \mathbb{R}^{\mathbb{R}} \times \mathbb{R}^{\mathbb{R}} \to \mathbb{R}^{\mathbb{R}}$ is the convolution operator.

Theorem 2.3. *For a continuous random vector $X : \Omega \to \mathbb{R}^m$ defined on the probability space $(\Omega, \mathcal{F}, P)$ with density $f_X : \mathbb{R}^m \to \mathbb{R}_+$ and an injective and smooth Borel measurable function $g : \mathbb{R}^m \to \mathbb{R}^m$, such that $Y = g(X)$ is a continuous random vector. Then the density of random vector Y is given by*

$$f_Y(y) = \frac{f_X(x)}{|J(y)|},$$

where $x = g^{-1}(y)$ and $J(y) = (J_{ij}(y) \triangleq \frac{\partial y_j}{\partial x_i} : i, j \in [m])$ is the Jacobian matrix.

Proof. For an injective map $g : \mathbb{R}^m \to \mathbb{R}^m$ we have $\{x\} = g^{-1}\{y\}$ for any $y \in g(\mathbb{R}^m)$. Further, since g is smooth, we have $dy = J(y)dx + o(|dx|)$, and thus

$$|dy| = |J(y)| \, |dx| + o(|dx|). \tag{1}$$

Defining set $dB(y) \triangleq \{w \in \mathbb{R}^m : y_j \leqslant w_j \leqslant y_j + dy_j\}$, we observe that for any continuous random vector $Y : \Omega \to \mathbb{R}^m$, we have

$$P \circ Y^{-1}(dB(y)) = f_Y(y) \, |dy| = P \circ X^{-1}\Big(dB(x)\Big) = f_X(x) \, |dx| \,. \tag{2}$$

We get the result by combining (**??**) and (**??**). $\square$

Example 2.4 (Sum of random variables). Suppose that $X : \Omega \to \mathbb{R}^2$ is a continuous random vector and $Y_1 = X_1 + X_2$. Let us compute $f_{Y_1}(y_1)$ using the above theorem. Let us define a random vector $Y : \Omega \to \mathbb{R}^2$ such that $Y = (X_1 + X_2, X_2)$ so that $|J(y)| = 1$. This implies, $f_Y(y) = f_X(x)$. Thus, we may compute the marginal density of Y_1 as,

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_X(x) \mathbb{1}_{\{x_2 = y_2, x_1 + x_2 = y_1\}} dy_2 = \int_{-\infty}^{\infty} f_X(y_1 - y_2, y_2) dy_2.$$

If X is an independent random vector, then

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1}(y_1 - y_2) f_{X_2}(y_2) dy_2 = (f_{X_1} * f_{X_2})(y_1),$$

where $*$ represents convolution.