In the second chapter, we're going to talk about **transformation** and **expectation**. In short, transformation induces new random variable with a known random variable and expectation is the functional a random variable and somehow provides a weighted average sense.

1 Transformation

Recall that random variable maps a specific probability space to a universal probability space: $(\mathcal{R}, \mathcal{B}, \mathcal{L})$ while preserving the measurablility. Now, we want to extend the idea and map $(\mathcal{R}, \mathcal{B})$ to itself!

1.1 Measurable function

Definition 1 (Borel measurable function) *Let* $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$, *then* g *is a Borel measurable function from* $(\mathcal{X}, \mathcal{B})$ *to* $(\mathcal{Y}, \mathcal{B})$ *if* $\forall B \in \mathcal{B}$ $g^{-1}(B) \in \mathcal{B}$.

Property 2 *If* $X : \Omega \to \mathcal{X}$ *is a random variable and* g *is a Borel measurable function from* $(\mathcal{X}, \mathcal{B})$ *to* $(\mathcal{Y}, \mathcal{B})$. *Then,* $g(X)$ *is a random variable from* Ω *to* $(\mathcal{Y}, \mathcal{B})$.

Proof: Simply check the sufficient condition of random variable. $\forall B \in \mathcal{B}$ consider

$$
\begin{aligned}
\{w : g(X(w)) \in B\} &= \{w : X(w) \in g^{-1}(B)\} \\
&= \{w : w \in X^{-1}[g^{-1}(B)]\} \\
&\in \mathcal{B}
\end{aligned}
$$

Since g is Borel measurable, $g^{-1}(B) \in \mathcal{B}$. Moreover, X is a random variable, thus $X^{-1}[g^{-1}(B)] \in \mathcal{B}$. ∎

If we go back to the construction of random variable X, we can see that what X really does is inducing a probability space from $(\Omega, \mathcal{A}, P)$ to $(\mathcal{X}, \mathcal{B}, \mu_X)$, where μ_X is the induced measure defined as $\mu_X(B) = P(X \in B)$, $\forall B \in \mathcal{B}$.

Now, let's turn to a transformation of random variable, $Y = g(X)$, it's clearly that Y also induces a new probability space $(\mathcal{Y}, \mathcal{B}, \mu_Y)$. Moreover, here we can have two ways to define the measure μ_Y:

$$
\begin{aligned}
\mu_Y(B) &= P(g(X) \in B) && \text{(from P)} \\
&= \mu_X(X \in g^{-1}(B)) && \text{(from } \mu_X)
\end{aligned}
$$

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (\mathcal{X}, \mathcal{B}, \mu_X) \xrightarrow{g} (\mathcal{Y}, \mathcal{B}, \mu_Y)$$

1.2   p.m.f., p.d.f., and cumulative distribution

In the very beginning, let's define some notations here. For convenience, from now on X denotes a random variable from $(\Omega, \mathcal{A}, P)$ to $(\mathcal{X}, \mathcal{B}, \mu_X)$. And g is a Borel measurable function from $(\mathcal{X}, \mathcal{B})$ to $(\mathcal{Y}, \mathcal{B})$ that induces a random variable $Y = g(X)$ with measure μ_Y.

As long as we have the definition of induced probability measure, we can further define the density function and cumulative distribution.

Property 3 *The cumulative distribution of $Y = g(X)$ is*

$$
\begin{aligned}
F_Y(y) &= P(g(X) \le y), \ \forall y \in \mathbb{R} \\
&= P(\{x \in \mathcal{X} : g(x) \le y\}) \\
&= \int_{\{x \in \mathcal{X}: g(x) \le y\}} f_X(x) dx
\end{aligned}
$$

Remark: Note that the integration form might not be directly for a discrete random variable. Here, we use the expectation of $g(X)$ to present a way to represent discrete random variable in an integration form:

$$E[g(X)] = \int_{x \in \mathcal{X}} g(x) dF_X(x)$$

Note that here we define $dF_X(x) = F_X(x^- + dx) - F_X(x^-)$. For continuous r.v., this is the density at point x. As to discrete r.v., when there's some mass on point x, then the value will have a jump (step) at that point. One can see that this representation is equivalent to sum over the point that have mass.

Now, we might want to have a more convenient way to calculate the cumulative distribution and density function. However, we have to find the preimage of each region and compute the probability of that region over the measure of X. And for most of the cases this is not quite easy. Consider the case that $g(x) = x^2$, $\forall y \ge 0$, $\{x \in \mathcal{X} : g(x) \le y\} = [-y, y]$. We cannot find a similar form $(\le y)$ in the context of X.

But this is not the end of the world, soon we can see that as g is monotone, the relation can be easily build up. Intuitively, the monotonicity preserves the direction and result in the convenience of calculation.

Property 4 *Suppose $g(y)$ is **strictly monotone** on $\mathcal{X}$, then*

$$
\begin{aligned}
F_Y(y) &= F_X(g^{-1}(y)) && , \ g \text{ is increasing} \\
F_Y(y) &= 1 - F_X(g^{-1}(y)) && , \ g \text{ is decreasing}
\end{aligned}
$$

Proof: Since we have the monotonicity, $\forall y \in \mathcal{Y}$

$$\{x \in \mathcal{X} : g(x) \le y\} = \{x \in \mathcal{X} : x \le g^{-1}(y)\} \qquad , \; g \text{ is increasing}$$
$$= \{x \in \mathcal{X} : x \ge g^{-1}(y)\} \qquad , \; g \text{ is decreasing}$$

Intuitively, monotonicity preserves the relative position of points in $\mathcal{X}$ and $\mathcal{Y}$.   ■

Property 5 *Suppose $g(y)$ is monotone, then*

$$f_Y(y) = f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|\mathbf{1}_Y(y)$$

Intuitively, monotonicity also gives us a sense of **scaling** from X to Y. We can regard the term $|\frac{df_X(g^{-1}(y))}{dy}|$ simply as a scaling factor which is the relative exchange in the domain of $\mathcal{X}$ w.r.t. the domain of $\mathcal{Y}$.