STAT 7200

Introduction to Advanced Probability
Lecture 14

Taylor R. Brown

1. Theory of Convergence I
    - Convergence Almost Surely
    - Convergence in Probability
    - Law of Large Number

2. Foundation of Probability II
    - Distribution

Distribution

- **Distribution** For a random variable X on probability triple $(\Omega, \mathcal{F}, \mathrm{P})$, its distribution (law) is a function μ defined on all Borel subsets $\mathcal{B}$:

$$\mu(B) = \mathrm{P}(X \in B) = \mathrm{P}(\{\omega : \omega \in X^{-1}(B)\})$$

- We can then verify that $(\mathrm{R}, \mathcal{B}, \mu)$ is a valid probability triple. This probability triple is sometimes called the probability triple induced by random variable X.
- Notation-wise, we may write μ as $\mathcal{L}(X)$ and use $X \sim \mu$ to represent that μ is the distribution of X.
- **Remark** Generally speaking, if we want to compare two random variables X and Y directly, we need to make sure that they are defined on the same probability triple. However, such requirement is not necessary when we compare the distributions of two random variables, as both $\mathcal{L}(X)$ and $\mathcal{L}(Y)$ can be regarded as probability measures over the same sample space R and $\sigma-$algebra $\mathcal{B}$.

CDF

- **CDF** We define the cumulative distribution function of a random variable X as $F_X(x) = \mathrm{P}(X \leq x)$.
- A CDF is a right-continuous, non-decreasing function, and $\lim_{x \to -\infty} F_X(x) = 0, \lim_{x \to \infty} F_X(x) = 1$
- **CDF specifies the Law (Prop. 6.0.2)**. It is clear that, for two random variables X and Y, if $\mathcal{L}(X) = \mathcal{L}(Y)$, then $F_X(x) = F_Y(x)$ for all $x \in \mathrm{R}$. On the other hand, by the uniqueness of extension theorem, if $F_X(x) = F_Y(x)$ for all $x \in \mathrm{R}$, then $\mathcal{L}(X) = \mathcal{L}(Y)$ as well.

Distribution and Expectation

- When we discuss expectation of a random variable X, we define the expectation over the probability triple $(\Omega, \mathcal{F}, \mathrm{P})$. The following theorem shows that, the corresponding expectation defined over the probability triple induced by X would be the same.

Theorem 1 (Change of Variable Theorem)

On a probability triple $(\Omega, \mathcal{F}, \mathrm{P})$, let X be random variable with distribution μ, and f be any Borel-measurable function from $\mathrm{R} \to \mathrm{R}$. Assuming that all necessary expectations exist, we have:

$$\mathrm{E}_{\mathrm{P}}[f(X)] := \int_{\Omega} f(X(\omega))\mathrm{P}(d\omega) = \int_{-\infty}^{\infty} f(t)\mu(dt) := \mathrm{E}_{\mu}[f(X)]$$

Proof of the Change of Variable Theorem

1) When f is an indicator function: $f = 1_B$ where B is a Borel set:
$\mathsf{E}_{\mathsf{P}}(f) = \int_\Omega f(X(\omega))\mathsf{P}(d\omega) = \int_\Omega 1_{X(\omega)\in B}\mathsf{P}(d\omega) = \mathsf{P}(X \in B)$
$\mathsf{E}_\mu(f) = \int_{-\infty}^{\infty} f(t)\mu(dt) = \int_{\mathsf{R}} 1_{t\in B}\mu(dt) = \mu(B) = \mathsf{P}(X \in B)$
2) For simple function f (a finite linear combination of indicator function), the theorem holds as the expectation is linear.
3) For non-negative functions f, we can always construct a sequence of non-negative simple functions $f_n \nearrow f$ (e.g. define $f_n(X(\omega)) = \Psi_n(f(X(\omega))))$. Then by the monotone convergence theorem:
$\mathsf{E}_{\mathsf{P}}(f) = \lim_n \mathsf{E}_{\mathsf{P}}(f_n) = \lim_n \mathsf{E}_\mu(f_n) = \mathsf{E}_\mu(f)$.
4) For general function f, we can write $f = f^+ - f^-$, then by the linearity of expectation
$\mathsf{E}_{\mathsf{P}}(f) = \mathsf{E}_{\mathsf{P}}(f^+) - \mathsf{E}_{\mathsf{P}}(f^-) = \mathsf{E}_\mu(f^+) - \mathsf{E}_\mu(f^-) = \mathsf{E}_\mu(f)$.

Distribution and Expectation

- Based on the previous theorem, the expectation of a function of random variable X is determined by the distribution of X. Thus, for two random variables X, Y, if $\mathcal{L}(X) = \mathcal{L}(Y)$, then $\mathrm{E}[f(X)] = \mathrm{E}[f(Y)]$ for any measurable function f (assuming the expectations exist).
- On the other hand, if $\mathrm{E}[f(X)] = \mathrm{E}[f(Y)]$ for any measurable function f, by setting $f = 1_B$ for any Borel set B, we have $\mathrm{P}(X \in B) = \mathrm{E}[f(X)] = \mathrm{E}[f(Y)] = \mathrm{P}(Y \in B)$, that is $\mathcal{L}(X) = \mathcal{L}(Y)$

Corollary 2 (6.1.3.)

For two random variables X and Y, $\mathcal{L}(X) = \mathcal{L}(Y)$ if and only if $\mathrm{E}[f(X)] = \mathrm{E}[f(Y)]$ for any measurable function f (assuming the expectations exist.

- Another useful result is: if $\mathrm{P}(X = Y) = 1$, we have $\mathcal{L}(X) = \mathcal{L}(Y)$, so $\mathrm{E}[f(X)] = \mathrm{E}[f(Y)]$ for any measurable function f.

Distribution: Point Mass

- If a random variable X equals a constant c with probability 1 ($\mathrm{P}(X = c) = 1$), the distribution of X is called a point mass distribution δ_c.
- For any Borel set B, $\delta_c(B) = \mathrm{P}(X \in B) = \mathrm{P}(c \in B) = 1_B(c)$, which equals 1 if $c \in B$ but 0 otherwise.
- The CDF of X, $F_X(x) = \delta_c((-\infty, x])$ equals 0 for $x < c$ and equals 1 for $x \geq c$.
- For any measurable function f, as $\mathrm{P}(X = c) = 1$, $\mathrm{E}[f(X)] = \mathrm{E}[f(c)] = f(c)$.

Distribution: Mixture Distribution

- Given a sequence of probability distribution $\{\mu_i\}$, we can define the mixture distribution as $\mu(B) = \sum_i \beta_i \mu_i(B)$ where $\{\beta_i\}$ is a sequence of non-negative constants summing to 1, and B is any Borel set.
- It is easy to verify that μ is a proper probability measure on R and the Borel σ-algebra $\mathcal{B}$. We have the following results regarding the expectation with respect to a mixture distribution:

Proposition 3 (6.2.1)

For the mixture distribution defined above, let f be any Borel measurable function and assuming all the necessary expectations exist, we have:

$$\mathrm{E}_\mu(f) = \int f d\mu = \sum_i \beta_i \int f d\mu_i = \sum_i \beta_i \mathrm{E}_{\mu_i}(f)$$

- **Proof** For $f = 1_B$, the above equality is equivalent to $\mu(B) = \sum_i \beta_i \mu_i(B)$, which is true by definition. The general case follows by the linearity and the MCT.

Distribution: Discrete Distributions

- Any discrete distribution can be viewed as the mixture distribution of at most countable point mass distribution. If X can take value from the set $\{a_1, a_2, \ldots\}$ and $\mathrm{P}(X = a_i) = p_i$, then the distribution of X can be represented as

$$\mathcal{L}(X) = \sum p_i \delta_{a_i}$$

- Thus, for any function f, $\mathrm{E}[f(X)] = \sum p_i \mathrm{E}_{\delta_{a_i}}[f(X)] = \sum p_i f(a_i)$.
- For instance, if $X \sim Bin(n, p)$, then $\mathcal{L}(X) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k$, and $\mathrm{E}[f(X)] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} f(k)$.

Distribution: Absolutely Continuous Distributions

- For Borel-measurable function $f \geq 0$ so that $\int_{-\infty}^{\infty} f(t)\lambda(dt) = 1$ (λ represents the (non/sigma-) finite Lebesque measure), we can define a distribution μ as:

$$\mu(B) = \int_{-\infty}^{\infty} f(t) 1_B \lambda(dt) = \int_B f(t)\lambda(dt),\ B \in \mathcal{B}$$

- Such a distribution is known as an absolutely continuous distribution, and we usually use notations such as $\mu(dt) = f(t)\lambda(dt)$ or $\frac{d\mu}{d\lambda} = f$. f is either called a Radon-Nikodym derivative or a density.

Distribution: Absolutely Continuous Distributions

Proposition 4

For an absolutely continuous distribution μ with density f, let g be any Borel measurable function; then we have:

$$\mathsf{E}_{\mu}(g) = \int_{-\infty}^{\infty} g(t)\mu(dt) = \int_{-\infty}^{\infty} g(t)f(t)\lambda(dt)$$

- **Proof** For $g = 1_B$, the equality holds by definition. The general case follows by the linearity and monotone convergence theorem.

Distribution: Practical Calculations and Examples

- The expectation of discrete random variables can be calculated as sums.
- The expectations of absolutely continuous random variables can be calculated using a Riemann integral. Thus, for absolutely continuous random variable X with density f:
  $\mathsf{E}_\mu(g(X)) = \int g(t)\mu(dt) = \int g(t)f(t)\lambda(dt) = \int_{-\infty}^{\infty} g(t)f(t)dt$
- Both are special cases of a Lebesgue integrals.
- Neither ac nor discrete: let $X \sim \mu = \frac{1}{2}\delta_1 + \frac{1}{2}\mu_N$ where μ_N is the distribution of standard normal $N(0,1)$, then
  $\mathsf{E}_\mu(X) = \frac{1}{2}1 + \frac{1}{2}0 = \frac{1}{2}$, $\mathsf{E}_\mu(X^2) = \frac{1}{2}1 + \frac{1}{2}1 = 1$ and
  $\mathsf{Var}_\mu(X) = \mathsf{E}_\mu(X^2) - [\mathsf{E}_\mu(X)]^2 = \frac{3}{4}$.