STAT 5010: Advanced Statistical Inference

Lecturer: Tony Sit
Scribe: Ali Choo and Hon Ku To

Lecture 2
14 Sep 2020

2 10 Ways of Viewing a Random Variable

Define probability space (χ, ϕ, p) where

χ : Sample space with element ω
ϕ : $\sigma-$algebra with element
P : Probability measure which assigns probabilities to elements of $\mathcal{A}$ which satisfy

(i) $0 \leq p(A) \leq 1$.

(ii) $p(\chi) = 1$

(iii) If the element are disjoint i.e. $A_i \bigcap A_j = \{\emptyset\}$ for $i \neq j$, then

$$P(\bigcup_{i=1}^{n} A_i) = P(A_1 \bigcup A_2 ... \bigcup A_n) = \sum_{i=1}^{n} P(A_i).$$

2.1 Way #1. Random Variable

A function $X : \chi \rightarrow R$ such that image $X^{-1}(B)$ of any Borel set or elements of $\mathcal{A}$ is called a random variable. A $p-$tipple of r.v's is called random vector.

2.2 Way #2. Distribution Function

Associated with a random vector X on $(\chi, \mathcal{A}, P)$ is a distribution function $d.f.$: $F(\chi) = F_{x_1,...x_p}(x_1, x_2, ..., x_p) = P(\omega : X_1(\omega) \leq x_1, ..., X_p(\omega) \leq \chi_p)$ Note that, F is right-continuous with left limits (RCLL) [or càdlàg as in "*continue à droite, limite à gauche*"].

2.3 Way #3. τ^{th} Quantile $(0 < \tau < 1)$

For any scalar r.v X with $d.f.$ F, the quantity $\theta(\tau) = F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}$, $\tau \in (0, 1)$ is called the τ^{th} quartile of X or F. Specifically for $\tau = 1/2, \theta(1/2)$: Median, $\theta(1/4)$: Lower quartile, $\theta(3/4)$:upper quartile.

2.4 Way #4. Density Function

If the $d.f.F$ is absolutely continues with respect to the measure μ then F has a density function w.r.t μ. We interested in case where μ' is the league measure in which case can write $F(\chi) = \int_{-\infty}^{x} f(t)dt$, $f(t) = F'(t) = \partial F(t)/\partial t$.

Theorem 1 (Radon-Nikodym)**.** *If a finite measure P is absolute continuous w.r.t. a σ finite measure μ, then there exists a non-negative measurable function f such that*

$$P(A) = \int_A f d\mu = \int f 1_A d\mu.$$

This specific function f is called the Radon-Nikodym derivative of P w.r.t. μ (the density of p w.r.t. μ) denoted as $f = dp/d\mu$.

2.5 Way #5. Expectation

$$\begin{aligned} E(X) &= \int X(\omega)dp(\omega) = \int_{-\infty}^{\infty} xdF(x) = \int_{-\infty}^{\infty} xf(x)\,dx \\ E(aX+bY) &= aE(X) + bE(Y). \end{aligned}$$

2.6 Way #6: Moments

The kth central moment of a random variable X is

$$\mu_k = E(\{X - E(X)\}^k), \quad k = 1, 2 \ldots$$

In particular, μ_k for the cases $k = 2, 3, 4$ are closely related to the variance, skewness and kurtosis of X respectively as follows:

$$\begin{aligned} var(X) &= \mu_2. \\ \text{Skewness}(X) &= \mu_3/\sigma^3 && \text{, which measures the symmetry of } X. \\ \text{Kurtosis}(X) &= \mu_4/\sigma^4 && \text{, which measures the peakedness and tail length of } X. \end{aligned}$$

2.7 Way #7: Moment Generating Function (MGF)

The moment generating function of X is

$$m_X(t) = E(e^{tX}) = \int e^{tX}\, dF(x), \quad t \in \mathbb{R}.$$

When $m_X(t)$ and its derivatives exist in some neighbourhood of 0, we have

$$E(X^k) = \underbrace{m_X^{(k)}(0)}_{\text{the } k\text{th derivative of } m_X \text{ with respect to } t}, \quad k = 0, 1, 2, \ldots$$

Properties:

1. $m_{\mu+\sigma X}(t) = e^{\mu t} m_X(\sigma t)$.
2. $m_{X+Y}(t) = m_X(t)m_Y(t)$ if X and Y are independent.

Illustration

Suppose we have a discrete random variable on $\{0, 1, 2, \ldots\}$ with $\text{pr}(X = j) = a_j$, where $\text{pr}(X = j)$ is the probability mass function of X.

Define the "generating function" of X as

$$g(z) = \sum_{j=0}^{\infty} a_j z^j.$$

Since $\sum_{j=0}^{\infty} a_j = 1$, $|g(z)| \leq \sum_{j=0}^{\infty} |a_j||z|^j \leq \sum_{j=0}^{\infty} a_j = 1$ for any $|z| \leq 1$.

Consider the following derivatives:

$$\begin{aligned}
g'(z) = a_1 + 2a_2 z + 3a_3 z^2 + \cdots &= \sum_{j=1}^{\infty} j a_j z^{j-1}, \\
g''(z) = 2a_2 + 6a_3 z + \cdots &= \sum_{j=2}^{\infty} j(j-1) a_j z^{j-2}, \\
&\vdots \\
g^{(k)}(z) &= \sum_{j=k}^{\infty} \binom{j}{k} k! a_j z^{j-k}.
\end{aligned}$$

Thus

$$g^{(k)}(0) = k! a_k \quad \text{or} \quad a_k = (k!)^{-1} g^{(k)}(0).$$

So, all the information about a_k's are "contained" within the function g and is made accessible by simply differentiating it (repeatedly) and evaluating it at 0.

This means that the distribution of a non-negative integer valued random variable is uniquely defined by its generating function.

Restricting the absolute value of X between 0 and 1 can be quite restrictive.

Write $E(z^X) = E(e^{-\lambda X}), 0 \leq \lambda < \infty$.

So in the previous case,

$$E(e^{-\lambda X}) = \sum_{j=0}^{\infty} a_j e^{-\lambda x_j} = \begin{cases} \text{(discrete case)} \sum_j p_j e^{-\lambda x_j}, \\ \text{(continuous case)} \int e^{-\lambda u} f(u)\, du, \end{cases}$$

where x_j's are all possible values of X.

This formulation is the Laplace transform of X.

Example (c.f. Casella and Berger (2002) E.g. 2.3.10: Non-unique Moments)

Consider two probability density functions given by

$$\begin{aligned}
f_1(x) &= \frac{1}{\sqrt{2\pi}x} e^{-(\log x)^2/2},\ 0 \leq x < \infty, \\
f_2(x) &= f_1(x)\{1 + \sin(2\pi \log x)\},\ 0 \leq x < \infty.
\end{aligned}$$

(f_1 is the probability density function of a lognormal distribution.)

It can be shown that if $X_1 \sim f_1(x)$,

$$E(X_1^r) = e^{r^2/2}, \quad r = 0, 1, \dots$$

Suppose $X_2 \sim f_2(x)$. We have

$$E(X_2^r) = \int_0^\infty x^r f_1(x)\{1 + \sin(2\pi \log x)\}\, dx = E(X_1^r) + \int_0^\infty x^r f_1(x) \sin(2\pi \log x)\, dx.$$

Consider the transformation: $y = \log x - r$. You can show that the transformed integral is an odd function over $(-\infty, \infty)$.

Hence $\int_0^\infty x^r f_1(x) \sin(2\pi \log x)\, dx = 0$ and $E(X_1^r) = E(X_2^r)$ for $r = 0, 1, \dots$

Even though X_1 and X_2 have the same moments for all r, their probability density functions are different.

2.8 Way #8: Characteristic functions

The characteristic function of X is

$$\phi_X(t) = E(e^{itX}) = \int e^{itx}\, dF(x),$$

where $i^2 = -1$, $e^{itx} = \cos(tx) + i \sin(tx)$.
For multivariate case,

$$\phi_X(t) = E(e^{it^T X}),$$

where $t = (t_1, \dots, t_p)^T$, $X = (X_1, \dots, X_p)^T$.
Existence: $|E(e^{itX})| \leq E|e^{itX}| = E|\cos(tX) + i\sin(tX)| = E(\{\cos^2(tX) + \sin^2(tX)\}^{1/2}) = 1$.
(Because $|a + ib|^2 = (a + ib)(a - ib) = a^2 + b^2$).
Inversion Formula (See, for example, Billingsley, 1995)

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)\, dt, \\ F_X(x) - F_X(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-ity}}{-it} \phi_X(t)\, dt \quad \text{for points of continuity of } F\ x \text{ and } y. \end{aligned}$$

The inversion formula provides a correspondence between F (or f) and ϕ.

Any characteristic function is bounded by 1 (shown above) and is a uniformly continuous function on $\mathbb{R}^{(p)}$. [Exercise]

Theorem 2 (Uniqueness)**.** *Let X and Y be random k-vectors.*

(i) If $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbb{R}^k$, then $F_X = F_Y$.

(ii) If $m_X(t) = m_Y(t) < \infty$ for all t in a neighbourhood of 0, then $F_X = F_Y$. (c.f. Casella and Berger, 2002 Theorem 2.3.11)

Proof. (i) For any $a = (a_1, \dots, a_k)^T \in \mathbb{R}^k$, $b = (b_1, \dots, b_k)^T \in \mathbb{R}^k$, and $(a, b] = (a_1, b_1] \times \cdots \times (a_k, b_k]$ satisfying pr_X(the boundary of $(a, b]) = 0$,

$$\Pr_X((a, b]) = \lim_{c \to \infty} \int_{-c}^{c} \cdots \int_{-c}^{c} \frac{\phi_X(t_1, \dots, t_k)}{(-1)^{k/2}(2\pi)^k} \prod_{j=1}^{k} \frac{e^{-it_j a_j} - e^{-it_j b_j}}{t_j}\, dt_j.$$

(ii) (See next lecture's note)

$\square$

References

Billingsley, P. (1995), *Probability and measure*, A Wiley-Interscience publication, Wiley, 3rd ed.

Casella, G. and Berger, R. L. (2002), *Statistical inference*, Pacific Grove, Calif.]: Duxbury/Thomson Learning, 2nd ed.