Probability and Statistics: From Classical to Bayesian

Week 3

19th April - 26th April 2021

1 Moment Generating Function

1.1 Moments and Central Moments

Moments: The quantity $\mu_k^{'} = \mathrm{E}[X^k]$ (if it exists) is called the k^{th} moment of X for $k \in \{1, 2, ..\}$

Central Moments: The quantity $\mu_k = E[(X - \mu)^k]$ (if it exists) is called the k^{th} central moment of X for k $\in \{1, 2, ...\}$.

Trivially you can easily see that $\mu_1^{'}$ = mean and $\mu_1 = 0$.(Linearity of expectation gives away the second part readily, remember our mean is just a constant)

1.2 Moment Generating Function

Define $M_X : A \to R$ by

$$M_X(t) = E[e^{tX}] \quad t \in A$$

We call M_X the moment generating function (mgf) of the random variable X. We say that the mgf of a random variable X exists, if there exists a positive real number a such that (-a, a) $\subset$ A (i.e., if $M_X(t) = E[e^{tX}]$ is finite in an interval containing 0).

The name **moment generating function** to the transform M_X is derived from the fact that M_X can be used to generate moments of random variable X.
Let X be a random variable with mgf M_X, which is finite on an interval (-a,a); for some a > 0 (i.e., mgf of X exists). Then, we have the following:

1. $\mu_r^{'} = E[X^r]$ is finite for each $r \in \{1, 2, ...\}$,
2. $\mu_r^{'} = E[X^r] = M_X^r(0)$, where $M_X^r(0) = \frac{d^r}{dt^r} M_X(t)$ at t=0, the r-th derivative of $M_X(t)$ at the point 0 for each r $\in \{1, 2, ...\}$, and
3. $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \mu_r^{'}$ with t $\in$ (-a, a).

Proposition 1. Under the notation and assumptions of the theorem define $\psi_X : (-a, a) -> R$ by $\psi_X(t) = ln(M_X(t))$; $t \in (-a, a)$. Then

$$\mu_1^{'} = E[X] = \psi_X^{(1)}(0) \text{ and } \mu_2 = Var(X) = \psi_X^{(2)}(0)$$

where $\psi_X^{(r)}(0)$ denotes the r-th derivative of ψ_X.

Proof: We have, for $t \in (-a, a)$

$$\psi_X^{(1)}(t) = \frac{M_X^1(t)}{M_X(t)} \text{ and } \psi_X^{(2)}(t) = \frac{M_X(t)M_X^{(2)}(t) - (M_X^{(1)}(t))^2}{(M_X(t))^2}$$

Just put t=0 in the equation and use property 2 to get the desired result. (Note: $M_X(0) = 1$, because $e^{tx} = 1$ for t=0, giving $M_X(0) = E[1] = 1$)

2 Change of Variable in $\mathbb{R}$

Let $h : \mathbb{R} \to \mathbb{R}$ be a function. Given the distribution of X, how will you find the distribution of $h(X)$?

1. CDF technique: The distribution $Z = h(X)$ can be determined by computing the distribution function.
Fix z $\in$ R,

$$F_Z(z) = P(Z \leq z) = P(h(X) \leq z).$$

Depending on the properties of the function h, we may, or may not be able to derive this probability in a closed form expression.

Example. Let X be a random variable with pmf

$$F_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{(n-x)}, & \text{if } x \in \{1, 2...n\} \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and p $\in$ (0, 1). Find the distribution function of Y = n - X. Note that $R_X = R_Y = \{0, 1, ...n\}$. For $y \in R_Y$, we get

$$P(Y \leq y) = P(X \geq n - y) = \sum_{x=n-y}^{n} \binom{n}{x} p^x (1-p)^{(n-x)} = \sum_{x=0}^{y} \binom{n}{n-x} p^{(n-x)} (1-p)^x$$

Thus, the distribution function of Y is

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ p^n & 0 \leq y < 1 \\ \sum_{j=0}^{i} \binom{n}{j} p^{(n-j)} (1-p)^j, & i \leq y < i + 1 \text{ for } i = 1, 2, ..., n - 1 \\ 1 & y \geq n \end{cases}$$

2.a. Change of variable (for discrete probability distributions): Let X be a random variable of discrete type with support S_X and pmf f_X. Define $Z = h(X)$.

In probability theory, the support of a probability distribution can be loosely thought of as the closure of the set of possible values of a random variable having that distribution.

Then, Z is a random variable of discrete type with support $S_Z = \{h(x) : x \in R_X\}$ with pmf

$$f_Z(z) = \begin{cases} \Sigma_{a \in A_z} f_X(x) & \text{if } z \text{ is in } S_Z \\ 0 & \text{otherwise} \end{cases}$$

where $A_z = \{\text{x} \in R_X : h(x) = z\}$.

If h is one-one then we can introduce $h^{-1}(x) : D \to \mathbb{R}$ where $D = \{h(x) : x \in \mathbb{R}\}$

$$f_Z(z) = \begin{cases} \text{f}_X(h^{-1}(z)) & \text{if } z \text{ is in } S_Z \\ 0 & \text{otherwise} \end{cases}$$

You will see example(s) regarding this in this week's written notes.

2.b. Change of variable (for continuous probability distributions): Let X be a continuous random variable with pdf f_X and support S_X. Let $h : \mathbb{R} \to \mathbb{R}$ be *differentiable and strictly monotone* on S_X (i.e., either $h^{'}(x) < 0$ for all $x \in S_X$ or $h^{'}(x) > 0$ for all $x \in S_X$). Let $S_T = \{h(x) : x \in S_X\}$. Basically we want h(x) to be invertible, the technique works for non-invertible functions but there are a lot of complications that come in with non-invertible functions, so lets skip that.
Then, T = h(X) is a continuous random variable with pdf

$$f_T(t) = \begin{cases} f_X(h^{-1}(t)) |\frac{d}{dt} h^{-1}(t)| & \text{if } t \text{ is in } S_T \\ 0 & \text{otherwise} \end{cases}$$

You will see example(s) regarding this in this week's written notes.

3 Joint distribution

Let $X_1, X_2, ..., X_m$ be random variables on the same probability space. We call X = $(X_1, \ldots, X_m)$ a random vector, as it is just a vector of random variables. The CDF of X, also called the joint CDF of $X_1, \ldots, X_m$ is the function $F_X : R^m \to [0, 1]$ defined as

$$F_X(t_1, \ldots, t_m) = P\{X_1 \le t_1, \ldots, X_m \le t_m\} = P\left\{ \bigcup_i^m \{X_i \le t_i\} \right\}$$

Properties of joint CDFs: The following properties of the joint CDF $F_X : R^m \to [0, 1]$ are analogous to those of the 1-dimensional CDF :

- F_X is increasing in each co-ordinate, i.e., if $s_1 \le t_1, \ldots, s_m \le t_m$, then $F_X(s_1, \ldots, s_m) \le F_X(t_1, ..., t_m)$.
- $\lim F_X(t_1, ..., t_m) = 0$ if $\max\{t_1, \ldots, t_m\} \to -\infty$ (i.e., one of the t_i goes to - ∞).
- $\lim F_X(t_1, \ldots, t_m) = 1$ if $\min\{t_1, \ldots, t_m\} \to +\infty$ (i.e., all of the t_i goes to $+\infty$).
- F_X is right continuous in each co-ordinate. That is $F_X(t_1+h_1, \ldots, t_m+h_m) \to F_X(t_1, \ldots, t_m)$ as $h_i \to 0^+$ for i = 1,. . . , m.

Joint pmf and pdf: Just like in the case of one random variable, we can consider the following two sub-classes of vector of random variables.

(1) Distributions with a pmf. These are CDFs for which there exist points $t_1, t_2, \ldots$ in $\mathbb{R}^m$ and non-negative numbers w_i such that $\Sigma_i w_i = 1$ (often we write $f(t_i)$ in place of w_i). For every t $\in \mathbb{R}^m$, we have

$$F(t) = \Sigma_{i:t_i \leq t} w_i$$

where s *leq* t means that each co-ordinate of s is less than, or equal to the corresponding co-ordinate of t.

(2) Distributions with a pdf. These are CDFs for which there is a non-negative function (may assume piecewise continuous for convenience) $f : R^m \to R_+$such that for every t $\in \mathbb{R}^m$ we have

$$F(t) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_n} f(u_1, \ldots, u_n)\, du_1 \ldots du_n$$

You can get PDF/CDF for any subset of co-ordinates from joint PDF/CDF. In case of pdfs, one can integrate out rest of the co-ordinates and in case of CDFs, put the ones not needed to $t_i \to \infty$.

We will continue the discussion on joint distributions next week, lets keep it till here for this week.