Probability and statistics

Lecture notes from April 2nd.

Definition 1. *The moment-generating function of the random variable X is defined as*

$$M_X(t) = \mathrm{E}\left(\mathrm{e}^{tX}\right). \tag{1}$$

In the discrete case,

$$M_X(t) = \sum_k \exp\left(tx_k\right) \cdot p_k,$$

while in the continuous case,

$$M_X(t) = \int_{\mathbb{R}} \mathrm{e}^{tx} \cdot f(x)\, dx.$$

Theorem 1. *Suppose the variables X, Y are independent. Let $V = aX + b$, $Z = X + Y$, where $a \neq 0$. Then $M_V(t) = \mathrm{e}^{\,tb} \cdot M_X(at)$ and $M_Z(t) = M_X(t) \cdot M_Y(t)$.*

Proof.

$$M_V(t) = \mathrm{E}\left(\mathrm{e}^{t(aX+b)}\right) = \mathrm{E}\left(\mathrm{e}^{tb}\,\mathrm{e}^{(at)X}\right) = \mathrm{e}^{tb} \cdot M_X(at),$$
$$M_Z(t) = \mathrm{E}\left(\mathrm{e}^{t(X+Y)}\right) = \mathrm{E}\left(\mathrm{e}^{tX}\,\mathrm{e}^{tY}\right) \overset{(a)}{=} \mathrm{E}\left(\mathrm{e}^{tX}\right) \cdot \mathrm{E}\left(\mathrm{e}^{tY}\right) = M_X(t) \cdot M_Y(t).$$

(a) Since the variables X, Y are independent, the variables e^{tX}, e^{tY} are also independent. □

Some examples of " MGFs "

Example:

(i) Let $X \sim B(n, p)$. Probabilities p_k are given by the formula $p_k = \binom{n}{k} p^k (1-p)^{n-k}$. Hence

$$M_X(t) = \sum_{k=0}^{n} \mathrm{e}^{tk}\, p_k = \sum_{k=0}^{n} \mathrm{e}^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(p \cdot \mathrm{e}^t\right)^k (1-p)^{n-k} \overset{(b)}{=} \left(p \cdot \mathrm{e}^t + q\right)^n.$$

(b) Binomial formula: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k}$.

(ii) Let $X \sim \text{Poisson}(\lambda)$. Now $p_k = \mathrm{e}^{-\lambda} \cdot \dfrac{\lambda^k}{k!}$. Since the expansion of e^x into an infinite series is (c): $\mathrm{e}^x = \sum_{k=0}^{\infty} \dfrac{x^k}{k!}$, it follows that $M_X(t) = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \dfrac{\lambda^k}{k!} \mathrm{e}^{tk} = \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \dfrac{\left(\lambda \cdot \mathrm{e}^t\right)^k}{k!} \overset{(c)}{=} \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}$.

(iii) The distribution $B(n, p)$ as a sum of the "$0-1$" distributions: We conduct n independent trials with probability of success p in each trial. If X_k is the random variable that counts the number of success in k attempts, then the distribution of X_k is called the "$0-1$" distribution.

X_k	0	1
p_k	$1-p$	p

For the moment generating function of the variable X_k, we have a simple expression

$$M_{X_k}(t) = \mathrm{E}\left(\mathrm{e}^{tX_k}\right) = (1-p) \cdot \mathrm{e}^0 + p \cdot \mathrm{e}^t = p \cdot \mathrm{e}^t + q.$$

Because $X = X_1 + \ldots + X_n$, and the variables X_k are independent, so[a]

$$M_X(t) = \prod_{k=1}^{n} M_{X_k}(t) = \left(p \cdot \mathrm{e}^t + q\right)^n.$$

(iv) Gamma distribution (b, p). Density $f(x) = \dfrac{b^p}{\Gamma(p)} x^{p-1} \mathrm{e}^{-bx}$, for $x \in (0, \infty)$, with parameters $b, p \in \mathbb{R}$.

$$M_X(t) = \frac{b^p}{\Gamma(p)} \int_0^\infty \mathrm{e}^{tx} x^{p-1} \mathrm{e}^{-bx}\, dx = \left\downarrow \begin{array}{rcl} u & = & (b-t)x \\ du & = & (b-t)\, dx \end{array} \right\uparrow = (*).$$

For $t \in (-\infty, b)$ we have (integration limits)

$$(*) = \frac{b^p}{\Gamma(p)} \cdot \frac{1}{(b-t)^p} \int_0^\infty \mathrm{e}^{-u} u^{p-1}\, du = \left(\frac{b}{b-t} \right)^p = \left(1 - \frac{t}{b} \right)^{-p}.$$

(v) We will now find the moment generating function of the distribution $N(0,1)$. The density of this distribution is $f(x) = \dfrac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-x^2/2}$, where $x \in \mathbb{R}$. The moment generating function is:

$$\begin{aligned} M_X(t) &= \mathrm{E}\left(\mathrm{e}^{tX}\right) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{tx} \mathrm{e}^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left[-\frac{x^2 - 2xt}{2} \right] dx = \\ &= \exp\left(\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left[-\frac{(x-t)^2}{2} \right] dx = \left| \begin{array}{l} u = x - t \\ du = dx \end{array} \right| = \\ &= \exp\left(\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-u^2/2}\, du = \mathrm{e}^{t^2/2}. \end{aligned}$$

[a] MGF of the sum of independent variables is the product of their MGFs.

Normal distribution

Definition 2. *We say that the random variable X has a normal distribution with parameters μ, σ^2 ($X \sim N(\mu, \sigma^2)$) only if the density is given by the formula*

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \text{ dla } x \in \mathbb{R}. \tag{2}$$

Theorem 2. *Let $X \sim N(\mu, \sigma^2)$ and let $Y = \dfrac{X - \mu}{\sigma}$, where $\sigma > 0$. Then $Y \sim N(0,1)$.*

Proof. We will prove the theorem in the classical way: from the cumulative distribution to density.

$$F_Y(t) = P(Y \le t) = P\left(\frac{X-\mu}{\sigma} \le t \right) = P(X \le \sigma t + \mu) = F_X(\sigma t + \mu).$$

By differentiating both sides of the above equation with respect to the variable t, we get

$$f_Y(t) = f_X(\sigma t + \mu) \cdot \sigma = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2} \right) \sim \mathrm{N}(0,1).$$

□

Definition 3. *Let $X_1, \ldots, X_n$ be independent random variables that follow the distribution $N(0,1)$. The distribution of the random variable $Z = \sum_{k=1}^{n} X_k^2$ is called the* chi-square distribution with n degrees of freedom *and we denote it by $Z \sim \chi^2(n)$.*

Conclusions (some of them will be content of the exercises)

1. ‘The converse of the theorem is as follows: If $S \sim \mathrm{N}(0,1)$ and $T = \sigma S + \mu$, then $T \sim \mathrm{N}(\mu, \sigma^2)$.

2. We know the moment generating function of the distribution $U \sim \mathrm{N}(0,1)$ is $M_U(t) = \exp\left(\frac{t^2}{2}\right)$. Using Theorem 1 (first formula) we get MGF of the distribution $\mathrm{N}(\mu, \sigma^2)$
$$M_U(t) = \exp\left(\mu\, t + \frac{\sigma^2 t^2}{2}\right).$$
3. If $U \sim \mathrm{N}(0,1)$, then $U^2 \sim \mathrm{Gamma}(1/2, 1/2) \equiv \chi^2(1)$.
4. Let $U_1, \dots, U_n$ be independent random variables with each having the distribution $\mathrm{N}(0,1)$. Also assume that, $Z = \sum_{k=1}^{n} U_k^2$. Then $Z \sim \mathrm{Gamma}(1/2, n/2) \equiv \chi^2(n)$.
5. Let $U_1, \dots, U_n$ be independent random variables with distributions (respectively) $\mathrm{N}(\mu_k, \sigma_k^2)$. Also assume that, $Z = \sum_{k=1}^{n} \alpha_k U_k$. Then $Z \sim \mathrm{N}\left(\sum_{k=1}^{n} \alpha_k \mu_k, \sum_{k=1}^{n} \alpha_k^2 \sigma_k^2\right)$.
6. If $U_1, \dots, U_n$ are independent random variables with each having the distribution $\mathrm{N}(\mu, \sigma^2)$, and $S^2 = \frac{1}{n} \cdot \sum_{k=1}^{n} (U_k - \mu)^2$, then $\frac{nS^2}{\sigma^2} \sim \chi^2(n)$.
7. $\mapsto \infty$. Other remarks...

Attention (with a normal 2-dimensional distribution)

Below, we give an informal description of the 2-dimensional normal distribution. Informal, because we can give a compact formula for the density of n-dimensional normal distribution (which will occur in the near future).

Let us consider the vector $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and the matrix $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. We also assume that the matrix Σ is non-singular(invertible). The density function $f(x,y)$ of the random variable (X,Y) is given as
$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right). \tag{3}$$
From the formula above, one can calculate that the marginal random variables X, Y have distributions $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. In addition, the covariance of the variables X, Y is $\rho\sigma_1\sigma_2$.

Three-dimensional normal distribution: the formula takes 3-4 lines. In summary: if we see the formula (3) on the general case, you can go past the inscription without any emotions, it is simply **normal** distribution.

Let's consider a 2-dimensional random variable $(X,Y) \sim \mathrm{N}(\mu, \Sigma)$. If the covariance $\mathrm{Cov}(X,Y) = 0$, then $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, or $\rho = 0$. If the formula (3) is rewritten as
$$f(x,y) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(x-\mu_1)^2}{\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\frac{(y-\mu_2)^2}{\sigma_2^2}\right). \tag{4}$$
then we conclude that the marginal variables X, Y are independent; Additionally $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$, and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. Therefore, we have a special case when the converse of Theorem 2 from note 4 holds, namely

Theorem 3. *The random variable $(X,Y) \sim N(\mu, \Sigma^2)$ is given. If $\mathrm{Cov}(X,Y) = 0$, then the marginal variables X, Y are independent.*

$\hookleftarrow$

Z poważaniem,
Witold Karczewski