

M378K: October 4th, 2024.

M378K Introduction to Mathematical Statistics

Problem Set #8

The Normal Distribution.

Definition 8.1. The moment-generating function (mgf) m_Y for a random variable Y is defined as

$$m_Y(t) = \mathbb{E}[e^{tY}]$$

for all t for which the above expectation exists. In fact, we say that the moment-generating function exists there exists a positive number b such that $m_Y(t)$ is finite for all t such that $|t| \leq b$.

Proposition 8.2. 1. If m_Y exists for a certain probability distribution, then it is unique.

2. If m_{Y_1} and m_{Y_2} are equal on an interval, then $Y_1 \stackrel{(d)}{=} Y_2$. identically distributed.

Corollary 8.3. Let Y_1 and Y_2 be independent and normally distributed. Define $Y = Y_1 + Y_2$. Then, the distribution of Y is ...

Proof. Note that $Y_i \sim N(\mu = \mu_i, \sigma_i)$ for $i = 1, 2$. Now, let's look at the mgf of Y . Then, since Y_1 and Y_2 are independent, we have

$$m_Y(t) = m_{Y_1}(t)m_{Y_2}(t).$$

We can now use the fact that for any $X \sim N(\mu, \sigma)$,

$$m_X(t) = e^{\mu t} m_Z(\sigma t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Hence,

$$m_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \times e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} = e^{(\mu_1 + \mu_2)t + \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}$$

We can conclude that $Y \sim N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$. □

Problem 8.1. Two scales are used to measure the mass m of a precious stone. The first scale makes an error in measurement which we model by a normally distributed random variable X_1 with mean $\mu_1 = 0$ and standard deviation $\sigma_1 = 0.04m$. The second scale is more accurate. We model its error by a normal random variable X_2 with mean $\mu_2 = 0$ and standard deviation $\sigma_2 = 0.03m$.

We assume that the measurements made using the two different scales are independent, i.e., that the random variables X_1 and X_2 are independent.

To get our final estimate of the mass of the stone, we take the average of the two results from the two different scales, i.e., we define $Y = \frac{X_1 + X_2}{2}$.

(i) What is the distribution of the random variable Y ? State the **name** of its distribution and the **values** of the parameters.

(ii) What is the probability that the error Y we get is within $0.005m$ of the actual mass of the stone? Namely, calculate

$$\mathbb{P}[|Y| < 0.005m].$$

$$Y \sim \text{Normal}(\mu_Y = 0, \sigma^2 = ?)$$

$$\begin{aligned} \text{Var}[Y] &= \text{Var}\left[\frac{1}{2}(X_1 + X_2)\right] \stackrel{\text{independence}}{=} \frac{1}{4} (\text{Var}[X_1] + \text{Var}[X_2]) \\ &= \frac{1}{4} (0.04^2 m^2 + 0.03^2 m^2) \\ &= \frac{1}{4} (0.05^2) m^2 = \left(\frac{0.05m}{2}\right)^2 \end{aligned}$$

$$Y \sim \text{Normal}(\mu_Y = 0, \sigma_Y = 0.025m)$$

$$\begin{aligned} \mathbb{P}[|Y| < 0.005m] &= \cancel{X} = \mathbb{P}[-0.005m < Y < 0.005m] \\ &= \mathbb{P}[Y < 0.005m] - \mathbb{P}[Y < -0.005m] \end{aligned}$$

$$Y = Z \cdot 0.025m$$

$$\begin{aligned} &= \mathbb{P}[Z \cdot 0.025m < 0.005m] - \mathbb{P}[Z \cdot 0.025m < -0.005m] \\ &= \mathbb{P}[Z < 0.2] - \mathbb{P}[Z < -0.2] \\ &= \Phi(0.2) - \Phi(-0.2) = 0.5793 - 0.4207 \approx 0.16 \end{aligned}$$

□

Y continuous

$$\mathbb{P}[Y \leq y] = \mathbb{P}[Y < y] + \mathbb{P}[Y = y]$$

Corollary 8.4. Let Y_1, \dots, Y_n be independent and identically distributed. Assume that $Y_1 \sim N(\mu, \sigma)$. Define

$$S = Y_1 + Y_2 + \dots + Y_n$$

Then, the distribution of S is ...

Proof. □

$$m_S(t) = m_{Y_1}(t) \cdot m_{Y_2}(t) \dots m_{Y_n}(t)$$

$$= (m_{Y_1}(t))^n$$

$$= \left(e^{\mu t + \frac{\sigma^2 t^2}{2}} \right)^n = e^{\underbrace{\mu}_{=\mu_S} n t + \underbrace{\sigma^2}_{=\sigma_S^2} \frac{n}{2} t^2}$$

$$\bar{Y} = \frac{1}{n} (Y_1 + \dots + Y_n)$$

$$S \sim N(n \cdot \mu, \sqrt{n} \cdot \sigma)$$

$$\Rightarrow \bar{Y} \sim \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Def'n. We say that Y_1 and Y_2 are identically distributed if

$$F_{Y_1}(y) = F_{Y_2}(y) \quad \text{for all } y \in \mathbb{R}$$

$$Y_1 \stackrel{(d)}{=} Y_2$$

$$tY_1 \stackrel{(d)}{=} tY_2$$

$$e^{tY_1} \stackrel{(d)}{=} e^{tY_2}$$

$$m_{Y_1}(t) = \mathbb{E}[e^{tY_1}] = \mathbb{E}[e^{tY_2}] = m_{Y_2}(t)$$

M378K Introduction to Mathematical Statistics

Problem Set #9

De Moivre-Laplace.

Problem 9.1. You are given a TRUE/FALSE exam with 30 questions. Suppose that you need to answer 21 questions correctly in order to pass. You have no idea what the class is about and decide to toss a fair coin to answer all the questions; you circle TRUE if the outcome is tails and you circle FALSE if the outcome is heads. What is your approximation of the probability p that you manage to pass the exam using this strategy?

$$N \sim b(n=30, p=1/2) \quad \text{exact: } \sum_{k=21}^{30} \binom{30}{k} \left(\frac{1}{2}\right)^{30}$$

For $Y \sim \text{Binomial}(n, p)$ we know that its probability mass function is:

$$p_Y(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \dots, n$$

Moreover, its expectation and its variance are

$$\mathbb{E}[Y] = np \quad \text{and} \quad \text{Var}[Y] = np(1-p).$$

Now, consider a sequence of binomial random variables $Y_n \sim \text{Binomial}(n, p)$. Note that, while the number of trials n varies, the probability of success in every trial p remains the same for all n . The normal approximation to the binomial is a theorem which states that

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0, 1) \quad \text{as } n \rightarrow \infty.$$

Practically, this means that Y_n is "approximately" normal with mean np and variance $np(1-p)$ for "large" n . The usual rule of thumb is that both $np > 10$ and $n(1-p) > 10$.

Another practical adjustment needs to be made due to the fact that discrete distributions of Y_n are approximated by a continuous (normal) distribution. This adjustment is usually referred to as the continuity correction. More specifically, provided that the conditions above are satisfied, for every integer $a < b$, we have that

$$\begin{aligned} \mathbb{P}[a \leq Y_n \leq b] &= \mathbb{P}\left[a - \frac{1}{2} < Y_n < b + \frac{1}{2}\right] \\ &= \mathbb{P}\left[\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}} < \frac{Y_n - np}{\sqrt{np(1-p)}} < \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right] \\ &\approx \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

where Φ , as usual, stands for the cumulative distribution function of the standard normal distribution.

For more about the history of the theorem and ideas for its proof, go to:

Wikipedia: [de Moivre-Laplace](#).

→: $N \sim b(30, 1/2)$

$$\mathbb{E}[N] = 15 ; \text{Var}[N] = 7.5 \Rightarrow \text{SD}[N] = \sqrt{7.5}$$

$$\mathbb{P}[N \geq 21] = \mathbb{P}[N \geq 20.5] = \mathbb{P}\left[\frac{N-15}{\sqrt{7.5}} \geq \frac{20.5-15}{\sqrt{7.5}}\right]$$

↑
cont.
corr.

$\overset{22}{N(0,1)}$

$$np = n(1-p) = 15 > 10$$

⋮