

1. Suppose X_1, \dots, X_n is a random sample from a distribution with mean μ and variance σ^2 . Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ be the sample mean and variance.

- (i) Find $E(\bar{X})$ and $\text{var}(\bar{X})$.
 (ii) Using $\sum (X_i - \bar{X})^2 = \sum \{(X_i - \mu) + (\mu - \bar{X})\}^2$ show that

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

By taking expectations show that $E(S^2) = \sigma^2$.

2. Let X_1, \dots, X_n be independent identically distributed random variables. Find the maximum likelihood estimators of the parameter θ for the following distributions. (In each case r is a known positive integer.)

- (i) X_i has a binomial distribution with parameters r and θ .
 (ii) X_i has a negative binomial distribution with probability mass function

$$f(x; \theta) = \binom{r+x-1}{x} \theta^r (1-\theta)^x, \quad x = 0, 1, 2, \dots$$

- (iii) X_i has a gamma distribution with probability density function

$$f(x; \theta) = \frac{\theta^r}{(r-1)!} x^{r-1} e^{-\theta x}, \quad x > 0.$$

3. Suppose that in a population of twins, males (M) and females (F) are equally likely to occur and that the probability that twins are identical is θ . If twins are not identical, their genders are independent (if they are identical, their genders are the same).

- (i) Show that, ignoring birth order, $P(MM) = P(FF) = (1+\theta)/4$ and $P(MF) = (1-\theta)/2$.
 (ii) Suppose that n twins are sampled. It is found that n_1 are MM , n_2 are FF , and n_3 are MF , but it is not known which twins are identical. Find the maximum likelihood estimator of θ .

4. Suppose X is a normal random variable with mean μ and variance σ^2 .

- (i) If a and b are constants, show that $aX + b$ has a normal distribution and find its mean and variance.
 (ii) If $Z = (X - \mu)/\sigma$, deduce that $Z \sim N(0, 1)$.
 (iii) Using (ii) find $P(X < x)$ in terms of Φ , where Φ is the cumulative distribution function of a $N(0, 1)$ random variable.
 (iv) If $c > 0$ is a constant, show that $P(\mu - c\sigma < X < \mu + c\sigma)$ does not depend on μ or σ .

5. It is a standard result, which you may assume, that if X_1 and X_2 are independent and normally distributed random variables, then $X_1 + X_2$ is normally distributed.

Suppose X_1, \dots, X_n are independent normal random variables, X_i having mean μ_i and variance σ_i^2 . If a_1, \dots, a_n are constants, show that $\sum_{i=1}^n a_i X_i$ is normally distributed and find its mean and variance.

6. In the previous question, you might find it frustrating to be told: “It is a standard result, which you may assume ...”. One nice way to prove the result is to use moment generating functions – see *Probability in 2nd year*. Here are the steps of a different proof.

(a) Let X and Y be independent $N(0, 1)$ random variables.

(i) Write down the joint density function $f_{X,Y}(x, y)$.

(ii) Let a and b be constants (not both zero). For any $z \in \mathbb{R}$, we know that

$$P(aX + bY \leq z) = \iint_A f_{X,Y}(x, y) dx dy$$

where A is the region of the xy -plane in which $ax + by \leq z$. By changing variables in this integral from (x, y) to (u, v) where $u = ax + by$, $v = bx - ay$, show that

$$P(aX + bY \leq z) = \int_{-\infty}^z \int_{-\infty}^{\infty} \frac{1}{2\pi(a^2 + b^2)} \exp\left[\frac{-(u^2 + v^2)}{2(a^2 + b^2)}\right] dv du.$$

(iii) Hence show that

$$P(aX + bY \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi(a^2 + b^2)}} \exp\left[\frac{-u^2}{2(a^2 + b^2)}\right] du.$$

[Remember that p.d.f.s integrate to 1, there's no need to actually do any integration.]

(iv) Deduce that $aX + bY \sim N(0, a^2 + b^2)$.

(b) Now suppose X_1 and X_2 are independent, $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2$. If a_1 and a_2 are constants (not both zero), use (a)(iv) (and question 4) to show that

$$a_1X_1 + a_2X_2 \sim N(a_1\mu_1 + a_2\mu_2, a_1^2\sigma_1^2 + a_2^2\sigma_2^2).$$