Probability and Statistics

(November- 8)

$X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

(In real life, we observe the realizations of these r.v.s also called as random sample or data)

We observe data:

We know the data/ random sample is from $N(\mu, \sigma^2)$.

We do not know the parameters μ and σ^2.

Objective: Guess / Estimate μ and σ^2.

Ex: $\bar{x}$ is an unbiased estimator of μ.

In fact, $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$(E(\bar{x}) = \mu)$$

$$var(\bar{x}) = \frac{\sigma^2}{n}$$

other examples of unbiased estimators of μ

$T = \alpha_1 x_1 + \cdots + \alpha_n x_n$ where $\alpha_i \in [0,1]$

$\Sigma \alpha_i = 1$

$$E(T) = E(\alpha_1 x_1 + \cdots + \alpha_n x_n)$$

$$= (\alpha_1 + \alpha_2 + \cdots + \alpha_n)\mu$$

$E(T) = \mu$ $\Rightarrow$ T is an unbiased estimator of μ.

Ex: Let $-3, -2, 0, 1, 2, 1.7, 2.9, -3.2, -1.6, 1.8, -1.9, 0.3, 0.01$ be a random sample from $N(\mu, 1)$

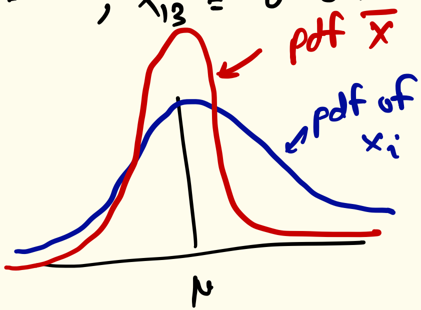
Estimate μ from this data??

$$\bar{x} = \frac{\sum_{i=1} x_i}{n} = \qquad \boxed{n = 13}$$

observed:

$x_1 = -3, \; x_2 = -2, \; x_3 = 0, \; x_4 = 1, \; \cdots, \; x_{13} = 0.01$

$$\bar{x} \sim N\left(\mu, \frac{1}{13}\right)$$



Observe:

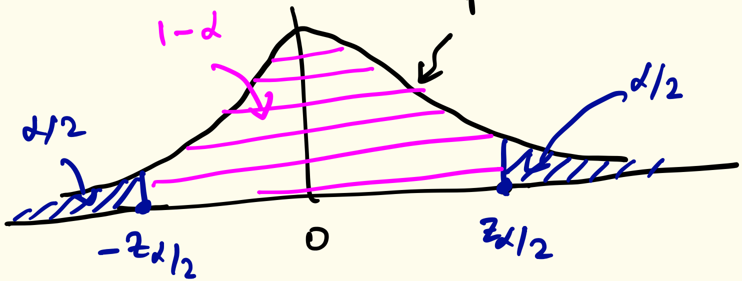
$$X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

$$\boxed{\bar{X} \sim N(\mu, \sigma^2/n)}$$

$$\Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\Rightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

pdf of $Z = N(0,1)$



$\alpha \in [0,1]$

$\frac{\alpha}{2} \in (0,1)$

$z_{\alpha/2} \in \mathbb{R}$

$P(Z \geq z_{\alpha/2}) = \alpha/2$

$$P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$$

$$\Rightarrow P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1-\alpha$$

Next objective:

let $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$X_1, \dots, X_n$ random sample from $N(\mu, \sigma^2)$

unknown parameters: μ, σ^2

objective: Estimate σ^2

choices for estimator: i) $T = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ biased estimator

ii) $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ ← unbiased estimator.

$X_1, X_2, \ldots, X_n$ sample of size 'n' from $N(\mu, \sigma^2)$

unknown: μ and σ^2.

objective: To estimate σ^2.

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \text{where } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

(sample mean)

To prove: s^2 is an unbiased estimator of σ^2.

i.e. $E(s^2) = \sigma^2$

Consider: $$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} \left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n}x_i$$

$$= \sum_{i=1}^{n}x_i^2 + n\bar{x}^2 - 2\bar{x}(n\bar{x})$$

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}x_i^2 - n\bar{x}^2$$

$$E\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right) = \sum_{i=1}^{n}E(x_i^2) - nE(\bar{x}^2)$$

Recall, $x_1, \ldots, x_n \overset{iid}{\sim} N(\mu, \sigma^2)$; $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$$E(x_i^2) = var(x_i) + [E(x_i)]^2$$
$$= \sigma^2 + \mu^2$$

$$var(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$$
$$E(\bar{x}^2) = var(\bar{x}) + [E(\bar{x})]^2$$
$$= \frac{\sigma^2}{n} + \mu^2$$

$$E\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)$$

$$= n\sigma^2 - \sigma^2$$

$$= (n-1)\sigma^2$$

$$\Rightarrow E\Big(\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}\Big) = \sigma^2$$

$$\Rightarrow E(S^2) = \sigma^2$$

$\Rightarrow$ S^2 is an unbiased estimator of σ^2.

$$E(T) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = \frac{n-1}{n}\sigma^2$$

$\Rightarrow T$ is a biased estimator (underestimates σ^2)

Distribution of $\frac{(n-1)S^2}{\sigma^2}$ $\boxed{\text{chi} - \chi}$

To prove: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ (chi-squared with $n-1$ degrees of freedom)

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$ for $i = 1, 2, \ldots, n$ and they are independent.

Recall: $Z_i^2 \sim \text{gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$ ✓

(Remember if $Y_i \sim \text{gamma}(\alpha_i, \lambda)$ and Y_i are indep $\Rightarrow \sum_{i=1}^{n} Y_i \sim \text{gamma}\left(\sum_{i=1}^{n} \alpha_i, \lambda\right)$)

Additive property of gamma density implies

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \underbrace{\text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)}_{\chi^2_n}$$

Recall, $X \sim \text{gamma}(\alpha, \lambda)$

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha - 1} \; ; \quad x > 0$$

$$= 0 \quad \text{o.w}$$

Pdf of Y (which is $\chi^2_n = \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$)

$$\boxed{f_Y(y) = \frac{\left(\frac{1}{2}\right)^{n/2}}{\Gamma(n/2)} e^{-y/2} y^{\frac{n}{2} - 1} \quad \text{for } y \geq 0 \\ = 0 \quad \text{o.w.}}$$

Result: Sum of squares of 'n' independent standard normal random variables is a chi-squared random variable with 'n' degrees of freedom.

Remember: $\chi^2_n = \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$

If $X_1, \dots, X_n$ is a random sample from $N(\mu, \sigma^2)$

$$\Rightarrow \quad \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n = \text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

MGF of $Y \sim \chi^2_n$ $\quad M_Y(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{n/2}$

$$\boxed{M_Y(t) = \left(\frac{1}{1-2t}\right)^{n/2} \quad ; -\infty < t < \frac{1}{2}}$$

$$\boxed{\left(\frac{\lambda}{\lambda - t}\right)^{\alpha}}$$

mgf of gamma (α, λ)

Fact: $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad ; \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Then $\bar{X}$ and S^2 are independent r.v.s.

Proof of $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Recall: $$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2_n$$

$$\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2_1$$

Consider

$$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}\left(x_i-\bar{x}+\bar{x}-\mu\right)^2$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + \sum_{i=1}^{n}(\bar{x}-\mu)^2$$

$$+2(\bar{x}-\mu)\cancelto{0}{\sum_{i=1}^{n}(x_i-\bar{x})}$$

$$\boxed{\begin{aligned}&\Sigma(x_i-\bar{x})\\&=\Sigma x_i-n\bar{x}\\&=0\end{aligned}}$$

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)^2$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\sigma^2} + \frac{(\bar{x}-\mu)^2}{\sigma^2/n}$$

$$\Rightarrow \quad U = V + W$$

where

$$U = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \sim \chi^2_n$$

$$V = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} = \frac{(n-1) s^2}{\sigma^2}$$

$$W = \frac{(\bar{x} - \mu)^2}{\sigma^2/n} \sim \chi^2_1$$

and since s^2 and $\bar{x}$ are independent, V and W are independent.

$$M_U(t) = M_{V+W}(t)$$

$$= M_V(t)\, M_W(t)$$

where $M_U(t)$, $M_V(t)$ and $M_W(t)$ are MGfs of U, V, W resp.

$$\left(\frac{1}{1-2t}\right)^{n/2} = M_V(t) \left(\frac{1}{1-2t}\right)^{1/2}$$

$$\Rightarrow \quad M_V(t) = \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}$$

$$\Rightarrow \quad V \sim \chi^2_{n-1} = \text{gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

$$\Rightarrow \quad \boxed{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}}$$