Sampling Distributions

Let X_i be a measurement of an experiment E_i, $i=1,\dots,n$.

Suppose f_i is the pmf/pdf of X_i

Let $f_{1:n}(x_1,\dots,x_n)$ be the joint pmf/pdf of

$X_1,\dots,X_n$.

We say $X_1,\dots,X_n$ are <u>independent</u> if

$$f_{1:n}(x_1,\dots,x_n) = f_1(x_1)\cdot f_2(x_2)\cdots f_n(x_n)$$

We say $X_1,\dots,X_n$ are <u>identically distributed</u> if

$$f_1 = f_2 = \dots f_n.$$

If $X_1,\dots,X_n$ are <u>independent + identically distributed</u> ← don't underline (IID), we say $X_1,\dots,X_n$ is a <u>random sample of size n</u> from the common distribution characterized by $f(x)$.

Example: Roll a die n times.

Definition A statistic is a function of a random sample.

Examples

- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (sample mean)
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ (sample variance)
- $X_2 + X_3$ (no name; not very interesting)

Since statistics are functions of random variables, they are themselves RVs. The distribution of a statistic is called a sampling distribution.

~~Four~~ Three Theorems: Assume $X_1, \dots, X_n$ are independent, and X_i has pmf / pdf $f_i(x)$, $\mu_i = E[X_i]$, $\sigma_i^2 = Var(X_i)$, MGF $M_{X_i}(t)$.

1. Let $Y = u(X_1, \dots, X_n)$ and let $g(y)$ denote its pmf / pdf. In the discrete case,

$$E[Y] = \sum_y y\, g(y) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} u(x_1, \dots, x_n) f_1(x_1) \cdots f_n(x_n)$$

In the continuous case, replace Σ by $\int$.

2. Suppose u_i is a function of X_i. If $E[u_i(X_i)]$ exists for $i = 1, \dots, n$, then

$$E[u_1(X_1) \cdots u_n(X_n)] = E[u_1(X_1)] \cdots E[u_n(X_n)].$$

3. If $Y = \sum_{i=1}^n a_i X_i$ where $a_i \in \mathbb{R}$,

then $\mu_Y = \sum_{i=1}^n a_i \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$.

and $M_Y(t) = \prod_{i=1}^n M_i(a_i t)$

product

Proofs

1. No proof, but an example $(n=2)$

X_i = roll of a fair, foursided die, $i=1,2$.

$Y = u(X_1, X_2) = X_1 + X_2$.

Then $f_1(x) = \frac{1}{4}$, $x = 1,..,4$ and $f_2(x) = f_1(x)$.

What is $g(y)$?

y	2	3	4	5	6	7	8
$g(y)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Therefore

$$E[Y] = \sum y\, g(y) = 2\left(\tfrac{1}{16}\right) + 3\left(\tfrac{2}{16}\right) + \cdots + 8\left(\tfrac{1}{16}\right)$$
$$= 5.$$

On the other hand, $$E[X_1 + X_2] = \sum_{x_1=1}^{4} \sum_{x_2=1}^{4} (x_1+x_2)\cdot \frac{1}{4} \cdot \frac{1}{4}$$
$$= 5.$$

Or $$E[X_1 + X_2] = E[X_1] + E[X_2]$$
$$= \frac{5}{2} + \frac{5}{2} = 5$$

2. We proved last time for $n=2$.
3. Assume $n=2$.

$$\mu_Y = E[Y] = E\left[\sum a_i X_i\right] = \sum a_i E[X_i]$$

$$= \sum a_i \mu_i \quad \text{by linearity of } E \text{ operator}$$

$$\sigma_Y^2 = E\left[(Y-\mu_Y)^2\right]$$

$$= E\left[(a_1 X_1 + a_2 X_2 - a_1\mu_1 - a_2\mu_2)^2\right]$$

$$= E\left[\left(a_1(X_1-\mu_1) + a_2(X_2-\mu_2)\right)^2\right]$$

$$= E\left[a_1^2 (X_1-\mu_1)^2\right] + 2E\left[a_1 a_2 (X_1-\mu_1)(X_2-\mu_2)\right]$$

$$+ E\left[a_2^2 (X_2-\mu_2)^2\right]$$

$$= a_1^2\sigma_1^2 + 0 + a_2^2\sigma_2^2$$

since X_1, X_2 are independent,

Finally,

$$M_Y(t) = E\left[e^{Yt}\right] = E\left[e^{(\Sigma a_i X_i)t}\right]$$

$$= E\left[e^{(a_1 X_1 + \cdots a_n X_n)t}\right]$$

$$= E\left[e^{a_1 X_1 t}\right] \cdots E\left[e^{a_n X_n t}\right]$$

$$= E\left[e^{X_1 (a_1 t)}\right] \cdots E\left[e^{X_n (a_n t)}\right]$$

$$= M_1(a_1 t) \cdots M_n(a_n t)$$

$$= \prod_{i=1}^{n} M_i(a_i t).$$

Corollary: If $X_1, \ldots, X_n$ are a random sample with common MGF $M(t)$ then

(a) If $Y = \Sigma X_i$, then

$$M_Y(t) = (M(t))^n$$

(b) If $\bar{X} = \frac{1}{n} \Sigma X_i$, then

$$M_{\bar{X}}(t) = \left(M\left(\frac{t}{n}\right)\right)^n$$