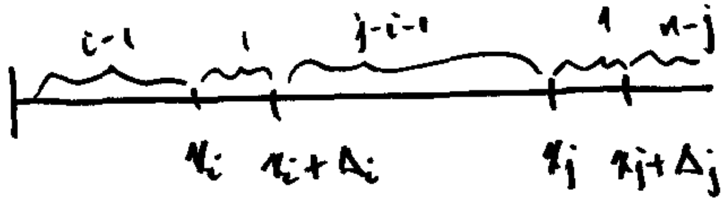
Stat 562
1-31-19

We were here: 1



Review of multinomial distribution

Run a sequence of n independent trials

Each trial has k possible outcomes with probabilities

$p_1, \ldots, p_k$, constant throughout the trials

$$\sum_{i=1}^{k} p_i = 1$$

2

let X_i = # of outcomes of type i

$$\sum_{i=1}^{k} X_i = n$$

The joint p.m.f. of $X_1, \ldots, X_k$ is

$$f(x_1, \ldots, x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!}\, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

$p_1 = F(x_i)$, $p_2 = F(x_i + \Delta_i) - F(x_i)$

$p_3 = F(x_j) - F(x_i + \Delta_i)$, $p_4 = F(x_j + \Delta_j) - F(x_j)$,

$p_5 = 1 - F(x_j + \Delta_j)$

③

$$c = \left(\frac{n!}{(i-1)!\,1!\,(j-i-1)!\,1!\,(n-j)!}\right) p_1^{i-1} p_2^{1} p_3^{j-i-1} p_4^{1} p_5^{n-j}$$

$$\underline{f_{ij}(x_i, x_j)} = \lim_{\substack{\Delta_i \to 0 \\ \Delta_j \to 0}} \frac{c}{\Delta_i \Delta_j}$$

$$= c \lim_{\substack{\Delta_i \to 0 \\ \Delta_j \to 0}} \left(F(x_i)\right)^{i-1} \left(\frac{F(x_i+\Delta_i)-F(x_i)}{\Delta_i}\right) \left(F(x_j)-F(x_i+\Delta_i)\right)^{j-i-1} \cdot \left(\frac{F(x_j+\Delta_j)-F(x_j)}{\Delta_j}\right)\left(1-F(x_j+\Delta_j)\right)^{n-j}$$

$$= \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} \left(F(x_i)\right)^{i-1} \left(F(x_j)-F(x_i)\right)^{j-i-1} \left(1-F(x_j)\right)^{n-j} f(x_i) \cdot f(x_j)$$

④

Example: $X_1, \ldots, X_n \sim iid$ Unif (a,b)

$$f(x) = \frac{1}{b-a} \quad a < x < b$$

Let $R = X_{(n)} - X_{(1)}$ Let $M = X_{(n)}$

Find the distribution of R, and find $E(R)$ & $V(R)$

$X_{(1)} = M - R \quad X_{(n)} = M$

$$J = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1$$

(5)

$$g(r,m) = f_{1n}(x_1, x_n)\,|J|$$

$$= \frac{n!}{(n-2)!}\left[F(x_n) - F(x_1)\right]^{n-2} f(x_1) f(x_n)$$

$$\boxed{F(x) = \frac{x-a}{b-a} \quad a<x<b}$$

$$g(r,m) = \frac{n!}{(n-2)!}\left(\frac{x_n-a}{b-a} - \frac{x_1-a}{b-a}\right)^{n-2} \frac{1}{b-a}\cdot\frac{1}{b-a}$$

$$= \frac{n!}{(n-2)!}\,\frac{1}{(b-a)^n}\Big(m-(m-r)\Big)^{n-2}$$

(6)

$$g(r,m) = \frac{n(n-1)\,r^{n-2}}{(b-a)^n}$$

$a < X_{(1)} < b$

$a < X_{(n)} < b$

$a < m-r < b, \; a < m < b$

$\Downarrow$

$r < m-a$

$r > m-b$

$X_{(n)} \geq X_{(1)}$

$m \geq m-r$

$$g(r) = \int_{r+a}^{b} \frac{n(n-1)r^{n-2}}{(b-a)^n}\,dm$$

$$= \frac{n(n-1)r^{n-2}}{(b-a)^n}\, m \Big|_{m=r+a}^{b} = \frac{n(n-1)r^{n-2}}{(b-a)^n}(b-a-r),$$

$0 \leq r < b-a$

⑦

Let $Y = \frac{R}{b-a}$ $\qquad r = (b-a)y$

$\frac{dr}{dy} = (b-a)$

$$h(y) = g(r)\left|\frac{dr}{dy}\right|$$

$$= \frac{n(n-1)r^{n-2}}{(b-a)^n}(b-a-r)(b-a)$$

$$= \frac{n(n-1)\cancel{(b-a)^{n-2}}\,y^{n-2}}{\cancel{(b-a)}^{n-1}}\left(\cancel{b-a} - \cancel{(b-a)}y\right)$$

$$= n(n-1)\,y^{n-2}(1-y) \qquad 0 < y < 1$$

⑧

This is the Beta density with $\alpha = n-1$, $\beta = 2$

$$E[Y] = \frac{\alpha}{\alpha+\beta} = \frac{n-1}{n+1}$$

$$V[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{(n-1)2}{(n+1)^2(n+2)}$$

$$R = (b-a)Y$$

$$\therefore\ E[R] = (b-a)\frac{n-1}{n+1}$$

$$V[R] = (b-a)^2\,\frac{2(n-1)}{(n+1)^2(n+2)}$$

(9)

Let X have a continuous distribution with cdf $F(x)$

<u>Defn</u>: The <u>p^{th} quantile</u> of X is

"ksi" $\xi_p = F^{-1}(p)$

Create the random variable $F(X_{(k)})$

Find $E[F(X_{(k)})]$

(10)

$$E[F(X_{(k)})] = \int_{-\infty}^{\infty} F(x_k)\, f_k(x_k)\, dx_k$$

Let $u = F(x_k)$

$du = f(x_k)\, dx_k$

$$E[F(X_{(k)})] = \int_0^1 u\, k \binom{n}{k} \underbrace{[F(x_k)]}_{u}{}^{k-1} \underbrace{[1-F(x_k)]}_{1-u}{}^{n-k} \underbrace{f(x_k)\, dx_k}_{du}$$

$$= k\binom{n}{k} \int_0^1 u^k (1-u)^{n-k}\, du$$

(11)

$$= k\binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)}\int_0^1 \underbrace{\frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)}u^k(1-u)^{n-k}}_{\text{Beta}(\alpha=k+1,\,\beta=n-k+1)}\,du$$

$$E[F(X_{(k)})] = k\binom{n}{k}\frac{k!\,(n-k)!}{(n+1)!}$$

$$= k\,\frac{n!}{\cancel{k!\,(n-k)!}}\,\frac{\cancel{k!\,(n-k)!}}{(n+1)!}$$

$$E[F(X_{(k)})] = \frac{k}{n+1}$$

(12)

That is, $F(X_{(k)})$ is an unbiased estimator of $\frac{k}{n+1}$

So $X_{(k)}$ is an estimator of $F^{-1}\left(\frac{k}{n+1}\right)$
(biased)

$$F^{-1}\left(\frac{k}{n+1}\right) = \xi_{\frac{k}{n+1}}$$

HW #4 due Feb 7

Chp 5 # 35, 39, 41

Stat 562 HW #4

5.35 Stirling's Formula (derived in Exercise 1.28), which gives an approximation for factorials, can be easily derived using the CLT.

(a) Argue that, if $X_i \sim \text{exponential}(1), i = 1, 2, \ldots$, all independent, then for every x,

$$P\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \leq x\right) \to P(Z \leq x),$$

where Z is a standard normal random variable.

(b) Show that differentiating both sides of the approximation in part (a) suggests

$$\frac{\sqrt{n}}{\Gamma(n)}(x\sqrt{n} + n)^{n-1}e^{-(x\sqrt{n}+n)} \approx \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

and that $x = 0$ gives Stirling's Formula.

5.39 This exercise, and the two following, will look at some of the mathematical details of convergence.

(a) Prove Theorem 5.5.4. (*Hint*: Since h is continuous, given $\varepsilon > 0$ we can find a δ such that $|h(x_n) - h(x)| < \varepsilon$ whenever $|x_n - x| < \delta$. Translate this into probability statements.)

(b) In Example 5.5.8, find a subsequence of the X_is that converges almost surely, that is, that converges pointwise.

5.41 Prove Theorem 5.5.13; that is, show that

$$P(|X_n - \mu| > \varepsilon) \to 0 \text{ for every } \varepsilon \quad \Leftrightarrow \quad P(X_n \leq x) \to \begin{cases} 0 & \text{if } x < \mu \\ 1 & \text{if } x \geq \mu. \end{cases}$$

(a) Set $\varepsilon = |x - \mu|$ and show that if $x > \mu$, then $P(X_n \leq x) \geq P(|X_n - \mu| \leq \varepsilon)$, while if $x < \mu$, then $P(X_n \leq x) \leq P(|X_n - \mu| \geq \varepsilon)$. Deduce the $\Rightarrow$ implication.

(b) Use the fact that $\{x : |x - \mu| > \varepsilon\} = \{x : x - \mu < -\varepsilon\} \cup \{x : x - \mu > \varepsilon\}$ to deduce the $\Leftarrow$ implication.

(See Billingsley 1995, Section 25, for a detailed treatment of the above results.)