Random Sampling and functions of random variables

Suppose, we have

$X_1, \ldots, X_n$ a random sample from a dist^n with p.d.f f_θ
or p.m.f. f_θ
(here $\theta \in \Theta$ is the characterising parameter)

"random sample" $\Rightarrow$ $X_1, \ldots, X_n$ are indep.

from the same dist^n $\Rightarrow$ $X_1, \ldots, X_n$ have identical dist^n

Thus "$X_1, \ldots, X_n$ is a random sample from f_θ"

$\Leftrightarrow$ "$X_1, \ldots, X_n$ are independently and identically distributed with f_θ"

Let Y = function of the random sample (not involving the parameter)

We call such functions "statistic"

e.g. $Y_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X} \rightarrow$ sample mean r.v.

$Y_2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = S_n^2 \rightarrow$ sample variance r.v.

or $(n-1)$

$Y_3 = \max(X_1, \ldots, X_n) = X_{(n)} \rightarrow$ maximum order statistic

$Y_4 = \min(X_1, \ldots, X_n) = X_{(1)} \rightarrow$ minimum order statistic

Point of interest: To know the prob law of such f^{n}s of the random sample

i.e. $(X_1, \ldots, X_n) \longrightarrow Y = f(X_1, \ldots, X_n)$
What is the p.d.f. / p.m.f. of Y

(I) M.g.f. based approach (provided m.g.f. exists)

Applicable for standard distributions with readily identifiable m.g.f.

Ex 1 : Additive property of standard dist^{n}s

(a) $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma^2)$

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

$$M_Y(t) = M_{\sum_1^n X_i}(t) = E\left(e^{t\sum X_i}\right)$$

$$= E\left(e^{tX_1} e^{tX_2} \cdots e^{tX_n}\right)$$

$$= \prod_{i=1}^{n} E\left(e^{tX_i}\right) \quad (\because X_1, \dots X_n \text{ are indep}).$$

$$= \prod_{i=1}^{n} e^{t\mu + \frac{t^2}{2}\sigma^2} = e^{tn\mu + \frac{t^2}{2} n\sigma^2}$$

$\Rightarrow Y \sim N(n\mu, n\sigma^2)$ by uniqueness of m.g.f.

If $X_1, \dots X_n$ are indep $N(\mu_i, \sigma_i^2)$, then $\sum_1^n X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$

(b) $X_1, \dots, X_n$ i.i.d. $P(\lambda)$

$Y = \sum_{i=1}^{n} X_i \sim P(n\lambda)$; $M_Y(t) = \prod_{i=1}^{n} e^{\lambda(e^t - 1)} = e^{n\lambda(e^t-1)}$

If $X_1, \dots, X_n$ are indep $P(\lambda_i)$

then $Y = \sum_1^n X_i \sim P\left(\sum_{i=1}^{n} \lambda_i\right)$.

(c) $X_1, \dots ; X_n$ ind $X_i \sim B(n_i, \underset{\text{same}}{p})$

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} \left(q + pe^t\right)^{n_i} = \left(q + pe^t\right)^{\sum n_i}$$

$$\Rightarrow Y \sim B\left(\sum_{i=1}^{n} n_i, p\right).$$

(II) Dist^n f^n based approach

Let $X_1, \ldots, X_n$ be indep continuous r.v.s with p.d.f. $f_X(x)$ and d.f. $F_X(x)$.

Let $Y = X_{(1)} = \min\{X_1, \ldots, X_n\}$ – smallest order statistic

$Z = X_{(n)} = \max\{X_1, \ldots, X_n\}$ – Largest order statistic

$$F_Y(y) = P(Y \le y) = P(\min\{X_1, \ldots, X_n\} \le y)$$

d.f. of Y

$$= 1 - P(\min\{X_1, \ldots, X_n\} > y)$$

$$= 1 - \prod_{i=1}^{n} P(X_i > y) \quad [\because \text{ of independence}]$$

$$= 1 - (1 - F_X(y))^n \quad [\because X_1, \ldots X_n \text{ have identical dist}^n]$$

(This step will hold for discrete setup also)

p.d.f. of Y:

$$f_Y(y) = n(1 - F_X(y))^{n-1} f_X(y)$$

Example: $X_1, \ldots, X_n$ i.i.d. (independent and identically distributed) $\exp(1)$

i.e. $f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & o/w \end{cases}$; $F_X(x) = 1 - e^{-x} \quad x > 0$

$$F_Y(y) = 1 - e^{-ny} \quad y > 0$$

$$f_Y(y) = \begin{cases} n e^{-ny} & y > 0 \\ 0, & o/w \end{cases}$$

$$Z = \max\{X_1, \ldots, X_n\}$$

$$F_Z(z) = P(Z \le z)$$

$$= P(\max\{X_1, \ldots, X_n\} \le z)$$

$$= P(X_1 \le z, \ldots, X_n \le z)$$

$$= \prod_{i=1}^{n} P(X_i \le z) \quad [\because \text{ of independence}]$$

$$= (F_X(z))^n \quad [\because \text{ of identical dist}^n]$$

(upto this step same for discrete & cont dist^n)

p.d.f. of Z:

$$f_Z(z) = n\,(F_X(z))^{n-1} f_X(z).$$

Example: $f_X(x) = e^{-x} \quad x > 0$; $F_X(x) = 1 - e^{-x} \quad x > 0$

$$F_Z(z) = (1 - e^{-z})^n$$

$$f_Z(z) = \begin{cases} n(1 - e^{-z})^{n-1} e^{-z}, & z > 0 \\ 0, & o/w. \end{cases}$$

Remark: In general, suppose jt p.d.f. of $(X_1, \ldots X_n)$ is

$$f_{X_1, \ldots X_n}(x_1, \ldots, x_n)$$

$$(X_1, \ldots, X_n) \to u(X_1, \ldots, X_n) = Y.$$

$$F_Y(y) = P(U(X_1, \ldots, X_n) \le y)$$

$$= \int \cdots \int_{u(x_1, \ldots x_n) \le y} f_{X_1 \ldots X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$