Useful Results - Statistics (4 pages; 26/5/20)

See also "Probability & Statistics - Important Ideas"

(1) Variance

(i) Sample variance $s^2 = \frac{1}{n-1}\left\{\left(\sum x_i{}^2\right) - n\overline{x}^2\right\}$

[assuming that it is to be used as an unbiased estimate for the population variance]

(ii) $Var(X) = E(X^2) - [E(X)]^2$

(2) Outliers

(i) To determine Q_1: take the items to the left of the median (or, if the median is the average of x_r & x_{r+1}, take the items up to and including x_r), and obtain their median. Similarly for Q_3.

[There are other methods, but exam mark schemes usually allow a certain amount of leeway, to cover all sensible methods.]

(ii) An outlier is defined as being less than $Q_1 - 1.5 \times IQR$ or greater than $Q_3 + 1.5 \times IQR$.

[An outlier is also sometimes defined as being more than 2 standard deviations from the mean.]

(3) Distributions

Discrete	
Uniform: $X\sim$ discrete $U(a,b)$	(i) $P(X=r)=\frac{1}{b-a+1}$ (ii) $E(X)=\frac{1}{2}(n+1)$ (iii) $Var(X)=\frac{1}{12}(n^2-1)$
Binomial: $X\sim B(n,p)$	pgf $G_X(s)=(q+ps)^n$
Geometric: $X\sim Geo(p)$ [X is no. of attempts needed for 1st success]	(i) $P(X=r)=q^{r-1}p$ (ii) $P(X\leq k)=1-q^k$ (iii) $E(X)=\frac{1}{p}$ (iv) $Var(X)=\frac{q}{p^2}$ (v) pgf $G_X(s)=\frac{ps}{1-qs}$
Negative Binomial [X is no. of attempts needed for n successes] [Becomes Geometric when $n=1$]	(i) prob. of nth success on rth attempt: $p_k=$ $\binom{r-1}{n-1}p^{n-1}q^{(r-1)-(n-1)}p$ $=\binom{r-1}{n-1}p^n q^{k-n}$ (ii) $E(X)=\frac{n}{p}$ (iii) $Var(X)=\frac{nq}{p^2}$ (iv) pgf $G_X(s)=\left(\frac{ps}{1-qs}\right)^n$
Poisson: $X\sim Po(\lambda)$	(i) $p_k=\frac{e^{-\lambda}\lambda^k}{k!}$ (ii) pgf $G_X(s)=e^{\lambda(s-1)}$
Continuous	
Uniform	(i) $f(x)=\frac{1}{b-a}$ (ii) $E(X)=\frac{1}{2}(a+b)$ (iii) $Var(X)=\frac{1}{12}(b-a)^2$
Normal: $X\sim N(\mu,\sigma^2)$	(i) $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$
Exponential	(i) $f(x)=\lambda e^{-\lambda x}$ (ii) $E(X)=\frac{1}{\lambda}$

[X is time between Poisson events]	(iii) $Var(X) = \frac{1}{\lambda^2}$

(4) Normal probabilities

sd	prob. (1 tail)
1	16%
1.645	5%
1.96	2.5%
2.326	1%
2.576	0.5%

(5) Correlation & Regression

(i) $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$

where $S_{xx} = \sum(x_i - \overline{x})^2 = \sum x_i{}^2 - n\bar{x}^2$ [& $S_{yy} = \sum(y_i - \overline{y})^2$ etc]

and $S_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y}) = \sum x_i y_i - n\bar{x}\bar{y}$,

(ii) $r_s = 1 - \frac{6\sum d_i{}^2}{n(n^2-1)}$

(iii) For the Regression line $y = a + bx$, $b = \frac{S_{xy}}{S_{xx}}$

(6) Probability Generating Functions

If $X_1, X_2, \ldots$ & N are independent random variables, where the X_i have pgf $G_X(s)$, then

(i) $S_N = X_1 + X_2 + \cdots + X_n$ has pgf $G_{S_N}(s) = G_N\left(G_X(s)\right)$

(ii) $E(S_N) = E(N)E(X)$

(iii) $Var(S_N) = E(N)Var(X) + Var(N)[E(X)]^2$

(7) de Morgan's Laws

$P[\,(A\cup B)'\,] = P[\,A'\cap B'\,]$

$P[\,(A\cap B)'\,] = P[\,A'\cup B'\,]$