1. Probabilities for events

For events A, B, and C $\qquad P(A \cup B) \;=\; P(A) + P(B) - P(A \cap B)$

More generally $\quad P(\bigcup A_i) \;=\; \sum P(A_i) - \sum P(A_i \cap A_j) + \sum P(A_i \cap A_j \cap A_k) - \cdots$

The odds in favour of A $\qquad P(A) \,/\, P(\overline{A})$

Conditional probability $\qquad P(A \,|\, B) \;=\; \dfrac{P(A \cap B)}{P(B)}$ provided that $P(B) > 0$

Chain rule $\qquad P(A \cap B \cap C) \;=\; P(A)\, P(B \,|\, A)\, P(C \,|\, A \cap B)$

Bayes' rule $\qquad P(A \,|\, B) \;=\; \dfrac{P(A)\, P(B \,|\, A)}{P(A)\, P(B \,|\, A) \;+\; P(\overline{A})\, P(B \,|\, \overline{A})}$

A and B are independent if $\qquad P(B \,|\, A) = P(B)$

A, B, and C are independent if $\qquad P(A \cap B \cap C) \;=\; P(A)P(B)P(C)\,,$ and

$$P(A \cap B) \;=\; P(A)P(B)\,, \quad P(B \cap C) \;=\; P(B)P(C)\,, \quad P(C \cap A) \;=\; P(C)P(A)$$

2. Probability distribution, expectation and variance

The probability distribution for a discrete random variable X is called the probability mass function (pmf) and is the complete set of probabilities $\{p_x\} = \{P(X = x)\}$

Expectation $\quad E(X) \;=\; \mu \;=\; \sum_x x p_x$

For function $g(x)$ of x, $\quad E\{g(X)\} = \sum_x g(x) p_x\,,\quad$ so $\quad E(X^2) = \sum_x x^2 p_x$

Sample mean $\quad \overline{x} \;=\; \dfrac{1}{n} \sum_k x_k \quad$ estimates μ from random sample $x_1, x_2, \ldots, x_n$

Variance $\quad \mathrm{var}\,(X) \;=\; \sigma^2 \;=\; E\{(X - \mu)^2\} \;=\; E(X^2) \;-\; \mu^2$

Sample variance $\quad s^2 \;=\; \dfrac{1}{n-1} \left\{ \sum_k x_k^2 \;-\; \dfrac{1}{n} \left(\sum_j x_j \right)^2 \right\} \quad$ estimates σ^2

Standard deviation $\quad \mathrm{sd}\,(X) \;=\; \sigma$

If value y is observed with frequency n_y

$$n = \sum_y n_y\,, \quad \sum_k x_k = \sum_y y n_y\,, \quad \sum_k x_k^2 = \sum_y y^2 n_y$$

Skewness $\quad \beta_1 \;=\; E \left(\dfrac{X - \mu}{\sigma} \right)^3 \qquad$ is estimated by $\quad \dfrac{1}{n-1} \sum \left(\dfrac{x_i - \overline{x}}{s} \right)^3$

Kurtosis $\quad \beta_2 \;=\; E \left(\dfrac{X - \mu}{\sigma} \right)^4 - 3 \qquad$ is estimated by $\quad \dfrac{1}{n-1} \sum \left(\dfrac{x_i - \overline{x}}{s} \right)^4 - 3$

Sample median $\quad \widetilde{x}$ or x_{med}. Half the sample values are smaller and half larger

If the sample values $x_1, \ldots, x_n$ are ordered as $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$, then $\widetilde{x} \;=\; x_{(\frac{n+1}{2})}$ if n is odd, and $\widetilde{x} \;=\; \frac{1}{2}\,(x_{(\frac{n}{2})} \;+\; x_{(\frac{n+2}{2})})$ if n is even

α-quantile $Q(\alpha)$ is such that $P(X \le Q(\alpha)) = \alpha$

Sample α-quantile $\widehat{Q}(\alpha)$ Proportion α of the data values are smaller

Lower quartile Q1 $= \widehat{Q}(0.25)$ one quarter are smaller

Upper quartile Q3 $= \widehat{Q}(0.75)$ three quarters are smaller

Sample median $\widetilde{x} = \widehat{Q}(0.5)$ estimates the population median $Q(0.5)$

3. Probability distribution for a continuous random variable

The cumulative distribution function (cdf) $F(x) = P(X \le x) = \int_{x_0=-\infty}^{x} f(x_0)\mathrm{d}x_0$

The probability density function (pdf) $f(x) = \dfrac{\mathrm{d}F(x)}{\mathrm{d}x}$

$E(X) = \mu = \int_{-\infty}^{\infty} x\, f(x)\mathrm{d}x\,,\quad \mathrm{var}\,(X) = \sigma^2 = E(X^2) - \mu^2,$ where $E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\mathrm{d}x$

4. Discrete probability distributions

Discrete Uniform *Uniform* (n)

$$p_x = \frac{1}{n} \quad (x = 1, 2, \ldots, n) \qquad \mu = (n+1)/2,\ \ \sigma^2 = (n^2-1)/12$$

Binomial distribution *Binomial* (n, θ)

$$p_x = \binom{n}{x} \theta^x (1-\theta)^{n-x} \quad (x = 0, 1, 2, \ldots, n) \qquad \mu = n\theta\,,\ \ \sigma^2 = n\theta(1-\theta)$$

Poisson distribution *Poisson* (λ)

$$p_x = \frac{\lambda^x e^{-\lambda}}{x!} \quad (x = 0, 1, 2, \ldots) \quad (\text{with } \lambda > 0) \qquad \mu = \lambda\,,\ \ \sigma^2 = \lambda$$

Geometric distribution *Geometric* (θ)

$$p_x = (1-\theta)^{x-1}\theta \quad (x = 1, 2, 3, \ldots) \qquad \mu = \frac{1}{\theta}\,,\ \ \sigma^2 = \frac{1-\theta}{\theta^2}$$

5. Continuous probability distributions

Uniform distribution *Uniform* (α, β)

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & (\alpha < x < \beta), \\ 0 & (\text{otherwise}). \end{cases} \qquad \mu = (\alpha+\beta)/2\,,\ \ \sigma^2 = (\beta-\alpha)^2/12$$

Exponential distribution *Exponential* (λ)

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (0 < x < \infty), \\ 0 & (-\infty < x \le 0). \end{cases} \qquad \mu = 1/\lambda\,,\ \ \sigma^2 = 1/\lambda^2$$

Normal distribution $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right\} \quad (-\infty < x < \infty), \quad E(X) = \mu, \quad \mathrm{var}(X) = \sigma^2$$

Standard normal distribution $N(0,1)$

If X is $N(\mu, \sigma^2)$, then $Y = \dfrac{X-\mu}{\sigma}$ is $N(0,1)$

6. Reliability

For a device in continuous operation with failure time random variable T having pdf $f(t)$ $(t > 0)$

The reliability function at time t	$R(t)$	$= P(T > t)$
The failure rate or hazard function	$h(t)$	$= f(t)/R(t)$
The cumulative hazard function	$H(t)$	$= \int_0^t h(t_0)\,\mathrm{d}t_0 = -\ln\{R(t)\}$
The Weibull(α, β) distribution has	$H(t)$	$= \beta t^{\alpha}$

7. System reliability

For a system of k devices, which operate independently, let

$R_i = P(D_i) = P($"device i operates"$)$

The system reliability, R, is the probability of a path of operating devices

A system of devices in series operates only if every device operates

$$R = P(D_1 \cap D_2 \cap \cdots \cap D_k) = R_1 R_2 \cdots R_k$$

A system of devices in parallel operates if any device operates

$$R = P(D_1 \cup D_2 \cup \cdots \cup D_k) = 1 - (1-R_1)(1-R_2)\cdots(1-R_k)$$

8. Covariance and correlation

The covariance of X and Y $\quad \mathrm{cov}(X,Y) = E(XY) - \{E(X)\}\{E(Y)\}$

From pairs of observations $(x_1, y_1), \ldots, (x_n, y_n)$ $\quad S_{xy} = \sum_k x_k y_k - \frac{1}{n}(\sum_i x_i)(\sum_j y_j)$

$$S_{xx} = \sum_k x_k^2 - \frac{1}{n}(\sum_i x_i)^2, \qquad S_{yy} = \sum_k y_k^2 - \frac{1}{n}(\sum_j y_j)^2$$

Sample covariance	$s_{xy} = \dfrac{1}{n-1} S_{xy}$	estimates $\mathrm{cov}(X,Y)$
Correlation coefficient	$\rho = \mathrm{corr}(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\mathrm{sd}(X)\cdot\mathrm{sd}(Y)}$	
Sample correlation coefficient	$r = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$	estimates ρ

9. Sums of random variables

$$E(X+Y) = E(X)+E(Y)$$
$$\mathrm{var}\,(X+Y) = \mathrm{var}\,(X)+\mathrm{var}\,(Y)+2\,\mathrm{cov}\,(X,Y)$$
$$\mathrm{cov}\,(aX+bY,\ cX+dY) = (ac)\,\mathrm{var}\,(X)+(bd)\,\mathrm{var}\,(Y)+(ad+bc)\,\mathrm{cov}\,(X,Y)$$

If X is $N(\mu_1,\sigma_1^2)$, Y is $N(\mu_2,\sigma_2^2)$, and $\mathrm{cov}\,(X,Y)=c$, then $X+Y$ is $N(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2+2c)$

10. Bias, standard error, mean square error

If t estimates θ (with random variable T giving t)

Bias of t $\quad \mathrm{bias}\,(t) = E(T)-\theta$

Standard error of t $\quad \mathrm{se}\,(t) = \mathrm{sd}\,(T)$

Mean square error of t $\quad \mathrm{MSE}(t) = E\{(T-\theta)^2\} = \{\mathrm{se}\,(t)\}^2+\{\mathrm{bias}\,(t)\}^2$

If $\overline{x}$ estimates μ, then $\mathrm{bias}\,(\overline{x})=0$, $\mathrm{se}\,(\overline{x})=\sigma/\sqrt{n}$, $\mathrm{MSE}(\overline{x})=\sigma^2/n$, $\widehat{\mathrm{se}}\,(\overline{x})=s/\sqrt{n}$

Central limit property If n is fairly large, $\overline{x}$ is from $N(\mu,\ \sigma^2/n)$ approximately

11. Likelihood

The likelihood is the joint probability as a function of the unknown parameter θ.

For a random sample $x_1,x_2,\ldots,x_n$

$$\ell(\theta;\, x_1,x_2,\ldots,x_n) = P(X_1=x_1\,|\,\theta)\ \cdots\ P(X_n=x_n\,|\,\theta) \qquad \text{(discrete distribution)}$$
$$\ell(\theta;\, x_1,x_2,\ldots,x_n) = f(x_1\,|\,\theta)\,f(x_2\,|\,\theta)\ \cdots\ f(x_n\,|\,\theta) \qquad \text{(continuous distribution)}$$

The maximum likelihood estimator (MLE) is $\widehat{\theta}$ for which the likelihood is a maximum

12. Confidence intervals

If $x_1,x_2,\ldots,x_n$ are a random sample from $N(\mu,\sigma^2)$ and σ^2 is known, then

the 95% confidence interval for μ is $(\overline{x}-1.96\frac{\sigma}{\sqrt{n}},\ \overline{x}+1.96\frac{\sigma}{\sqrt{n}})$

If σ^2 is estimated, then from the Student t table for t_{n-1} we find $t_0=t_{n-1,0.05}$

The 95% confidence interval for μ is $(\overline{x}-t_0\frac{s}{\sqrt{n}},\ \overline{x}+t_0\frac{s}{\sqrt{n}})$

13. Standard normal table — Values of pdf $\phi(y) = f(y)$ and cdf $\Phi(y) = F(y)$

y	$\phi(y)$	$\Phi(y)$	y	$\phi(y)$	$\Phi(y)$	y	$\phi(y)$	$\Phi(y)$	y	$\Phi(y)$
0	.399	.5	.9	.266	.816	1.8	.079	.964	2.8	.997
.1	.397	.540	1.0	.242	.841	1.9	.066	.971	3.0	.999
.2	.391	.579	1.1	.218	.864	2.0	.054	.977	0.841	.8
.3	.381	.618	1.2	.194	.885	2.1	.044	.982	1.282	.9
.4	.368	.655	1.3	.171	.903	2.2	.035	.986	1.645	.95
.5	.352	.691	1.4	.150	.919	2.3	.028	.989	1.96	.975
.6	.333	.726	1.5	.130	.933	2.4	.022	.992	2.326	.99
.7	.312	.758	1.6	.111	.945	2.5	.018	.994	2.576	.995
.8	.290	.788	1.7	.094	.955	2.6	.014	.995	3.09	.999

14. Student t table — Values $t_{m,p}$ of x for which $P(|X| > x) = p$, when X is t_m

m	$p=$ 0.10	0.05	0.02	0.01	m	$p=$ 0.10	0.05	0.02	0.01
1	6.31	12.71	31.82	63.66	9	1.83	2.26	2.82	3.25
2	2.92	4.30	6.96	9.92	10	1.81	2.23	2.76	3.17
3	2.35	3.18	4.54	5.84	12	1.78	2.18	2.68	3.05
4	2.13	2.78	3.75	4.60	15	1.75	2.13	2.60	2.95
5	2.02	2.57	3.36	4.03	20	1.72	2.09	2.53	2.85
6	1.94	2.45	3.14	3.71	25	1.71	2.06	2.48	2.78
7	1.89	2.36	3.00	3.50	40	1.68	2.02	2.42	2.70
8	1.86	2.31	2.90	3.36	∞	1.645	1.96	2.326	2.576

15. Chi-squared table — Values $\chi^2_{k,p}$ of x for which $P(X > x) = p$, when X is χ^2_k and $p =$ .995, .975, *etc*

k	.995	.975	.05	.025	.01	.005	k	.995	.975	.05	.025	.01	.005
1	.000	.001	3.84	5.02	6.63	7.88	18	6.26	8.23	28.87	31.53	34.81	37.16
2	.010	.051	5.99	7.38	9.21	10.60	20	7.43	9.59	31.42	34.17	37.57	40.00
3	.072	.216	7.81	9.35	11.34	12.84	22	8.64	10.98	33.92	36.78	40.29	42.80
4	.207	.484	9.49	11.14	13.28	14.86	24	9.89	12.40	36.42	39.36	42.98	45.56
5	.412	.831	11.07	12.83	15.09	16.75	26	11.16	13.84	38.89	41.92	45.64	48.29
6	.676	1.24	12.59	14.45	16.81	18.55	28	12.46	15.31	41.34	44.46	48.28	50.99
7	.990	1.69	14.07	16.01	18.48	20.28	30	13.79	16.79	43.77	46.98	50.89	53.67
8	1.34	2.18	15.51	17.53	20.09	21.95	40	20.71	24.43	55.76	59.34	63.69	66.77
9	1.73	2.70	16.92	19.02	21.67	23.59	50	27.99	32.36	67.50	71.41	76.15	79.49
10	2.16	3.25	13.31	20.48	23.21	25.19	60	35.53	40.48	79.08	83.30	88.38	91.95
12	3.07	4.40	21.03	23.34	26.22	28.30	70	43.28	48.76	90.53	95.02	100.4	104.2
14	4.07	5.63	23.68	26.12	29.14	31.32	80	51.17	57.15	101.9	106.6	112.3	116.3
16	5.14	6.91	26.30	28.85	32.00	34.27	100	67.33	74.22	124.3	129.6	135.8	140.2

16. The chi-squared goodness-of-fit test

The frequencies n_y are grouped so that the fitted frequency $\widehat{n}_y$ for every group exceeds about 5.

$X^2 = \sum_y \frac{(n_y - \widehat{n}_y)^2}{\widehat{n}_y}$ is referred to the table of χ^2_k with significance point p,

where k is the number of terms summed, less one for each constraint, *eg* matching total frequency, and matching $\overline{x}$ with μ

17. Joint probability distributions

Discrete distribution $\{p_{xy}\}$, where $p_{xy} = P(\{X = x\} \cap \{Y = y\})$.

Let $p_{x\bullet} = P(X = x)$, and $p_{\bullet y} = P(Y = y)$, then

$$p_{x\bullet} = \sum_y p_{xy} \quad \text{and} \quad P(X = x \,\big|\, Y = y) = \frac{p_{xy}}{p_{\bullet y}}$$

Continuous distribution

Joint cdf $\quad F(x,y) = P(\{X \le x\} \cap \{Y \le y\}) = \int_{x_0=-\infty}^{x} \int_{y_0=-\infty}^{y} f(x_0\,, y_0)\,\mathrm{d}x_0\,\mathrm{d}y_0$

Joint pdf $\quad f(x,y) = \frac{\mathrm{d}^2 F(x,y)}{\mathrm{d}x\,\mathrm{d}y}$

Marginal pdf of X $\quad f_X(x) = \int_{-\infty}^{\infty} f(x, y_0)\,\mathrm{d}y_0$

Conditional pdf of X given $Y = y$ $\quad f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ (provided $f_Y(y) > 0$)

18. Linear regression

To fit the linear regression model $y = \alpha + \beta x$ by $\widehat{y}_x = \widehat{\alpha} + \widehat{\beta}x$ from observations $(x_1, y_1), \ldots, (x_n, y_n)$, the least squares fit is $\widehat{\alpha} = \overline{y} - \overline{x}\widehat{\beta}$, $\widehat{\beta} = \frac{S_{xy}}{S_{xx}}$

The residual sum of squares $\quad \text{RSS} = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$

$\widehat{\sigma^2} = \frac{\text{RSS}}{n-2} \qquad \frac{n-2}{\sigma^2}\widehat{\sigma^2}$ is from χ^2_{n-2}

$$E(\widehat{\alpha}) = \alpha\,, \quad E(\widehat{\beta}) = \beta\,,$$

$$\operatorname{var}(\widehat{\alpha}) = \frac{\sum x_i^2}{n\,S_{xx}}\sigma^2\,, \quad \operatorname{var}(\widehat{\beta}) = \frac{\sigma^2}{S_{xx}}\,, \quad \operatorname{cov}(\widehat{\alpha}, \widehat{\beta}) = -\frac{\overline{x}}{S_{xx}}\sigma^2$$

$$\widehat{y}_x = \widehat{\alpha} + \widehat{\beta}x\,, \qquad E(\widehat{y}_x) = \alpha + \beta x\,, \qquad \operatorname{var}(\widehat{y}_x) = \left\{\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}\right\}\sigma^2$$

$\frac{\widehat{\alpha} - \alpha}{\widehat{\text{se}}\,(\widehat{\alpha})}\,, \quad \frac{\widehat{\beta} - \beta}{\widehat{\text{se}}\,(\widehat{\beta})}\,, \quad \frac{\widehat{y}_x - \alpha - \beta\,x}{\widehat{\text{se}}\,(\widehat{y}_x)}$ are each from t_{n-2}