1 General Probability Theory

- Normalisation of probability

$$\sum_i p_i \rightarrow \int_{-\infty}^{\infty} \rho(x) dx = 1 \tag{1}$$

- Continuous change of variables

$$\rho_y(y) = \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \rho_x(x), \tag{2}$$

with $\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$ generalising to Jacobian $|J|$ in higher dimensions.

- Expectation Value

$$E(x) = \langle x \rangle = \sum_i x_i P_i \rightarrow \int_{-\infty}^{\infty} x \rho(x) dx \tag{3}$$

- Variance

$$V(x) = \sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - x^2 \tag{4}$$

Standard deviation is $\sigma = \sqrt{V(x)}$.

- Independent events means joint probability factorises:

$$P(A, B) = P(A)P(B) \tag{5}$$

2 Discrete Distributions

2.1 Binomial

$$B(n; p, N) = \binom{N}{n} p^n (1-p)^{N-n} \tag{6}$$

with $E(n) = pN$, and $V(n) = Np(1-p)$.

2.2 Poisson

$$P(n; \mu) = \frac{\mu^n e^{-\mu}}{n!} \tag{7}$$

with $E(n) = \mu$ and $V(n) = \mu$.

- often, mean is calculated via rate $\mu = \lambda t$
- is limit of Binomial with $N \rightarrow \infty, p \rightarrow 0$, s.t. pN finite. Then,

$$N >> n, \qquad (1-p)^N \rightarrow e^{-pN}$$

.

3 Continuous Distributions

3.1 Uniform

$$U(x; a, b) = \begin{cases} \rho_0 & a \leq x \leq b \\ 0 & otherwise \end{cases} \tag{8}$$

where $\rho_0 = \frac{1}{b-a}$ is given by normalisation. Also $E(x) = \frac{b+a}{2}$ and $V(x) = \frac{(b-a)^2}{12}$.

3.2 Exponential Density Function

$$E(t; \lambda) = \lambda e^{-\lambda t} \tag{9}$$

with $E(t) = 1/\lambda$ and $V(t) = 1/\lambda^2$. Gives the distribution in time between consecutive Poisson distributed events.

3.3 Gaussian Density Function

$$G(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10}$$

with $E(x) = \mu$ and $V(x) = \sigma^2$. Usually in *standard score*: $Z = \frac{x-\mu}{\sigma}$.

4 Central Limit Theorem

The sum of N independent random variables will have a Gaussian PDF as N gets large, independent of the original distributions.

5 Likelihood

Say we believe to know the form of the PDF $P(x; \theta_j)$ depending on some parameters θ_j. Total probability of x_i independent measurements occurring with given set of parameters is the likelihood:

$$L(\theta_j; x_i) = \prod_{i=1}^{N} P(x_i; \theta_j). \tag{11}$$

To find best estimate for parameters, maximise log-likelihood

$$\frac{\partial \ln L}{\partial \theta_j} = 0, \qquad \forall j \tag{12}$$

as the logarithm is monotonously increasing.

5.1 Maximum Likelihood Principle

Parameters that maximise the log-likelihood are good best estimates.

5.2 Gaussian Approximation

The log-likelihood of Gaussian with known σ, estimating μ after one measurement x_1 is

$$\ln L(\mu) = -\ln \sqrt{2\pi}\sigma - \frac{(x_1 - \mu)^2}{2\sigma^2}, \tag{13}$$

which is the Taylor expansion of *any* function around a maximum. Hence, any PDF will look approx. Gaussian near the peak.

6 Least Squares

Define the residual as

$$r_i = y_i - f(x_i; \theta_j), \tag{14}$$

where f is the function supposed to fit the data y_i. Then, take sum of squares of residuals

$$S(\theta_j; y_i) = \sum_{i=1}^{N} r_i^2 \tag{15}$$

and minimise S to find parameters.

7 Chi-squared

Incorporate uncertainties σ_i by defining the pull as

$$p_i = \frac{r_i}{\sigma_i} \tag{16}$$

and minimise chi-squared

$$\chi^2(\theta_j; y_i) = \sum_{i=1}^{N} p_i^2 \tag{17}$$

to find parameters.

- If all σ_i are the same, gives same result as least squares.
- Have good fit if
$$\chi^2_{min} \sim N_{dof} = N_{data} - N_{parameters}$$
- use for binned distributions if all bins have large number of entries ($n_i > 20$)
- maximising $\ln L$ and minimising χ^2 are identical for Gaussian case as
$$\chi^2 \sim -2 \ln L$$

8 Parameter Uncertainties

8.1 Propagation of Errors Formula

$$\sigma_\theta^2 = \sum_i \left(\frac{\partial y_\theta}{\partial x_i}\right)^2 \sigma_{x_i}^2 \tag{18}$$

8.2 Maximum Likelihood

Gaussian approx:

$$L(\theta) \approx L(\hat{\theta}) e^{-\frac{(\theta-\hat{\theta})^2)}{2\Sigma^2}}, \tag{19}$$

Then the uncertainty Σ on our estimate $\hat{\theta}$ is given by

$$\frac{1}{\Sigma^2} = -\frac{\mathrm{d}^2 \ln L}{\mathrm{d}\theta^2}\Big|_{\hat{\theta}} \tag{20}$$

In general, (not only Gaussian), uncertainty is range that makes $\ln L$ decrease by the value 1/2.

8.3 Chi-squared

By Taylor expansion around min. (first derivative vanishes), can show

$$\frac{1}{\Sigma^2} = \frac{1}{2} \frac{\mathrm{d}^2 \chi^2}{\mathrm{d}\theta^2}\Big|_{\hat{\theta}} \tag{21}$$

In general, uncertainty is range in which χ^2 increases by one unit. As $\chi^2 \sim -2 \ln L$, both methods to determine the uncertainty of the estimated parameter θ give the same result.

9 Confidence Intervals

Define the confidence interval between (a, b), s.t.

$$\int_a^b \rho(x) dx = 68.3\%. \tag{22}$$

- Taking many measurements, 68.3% will fall into the confidence interval (frequentist).
- For a Gaussian, it is the range $\pm 1\sigma$ around the mean.
- Asymmetric distribution; can take two sided (i.e. 15.9% on each side) or one-sided distribution.
- Problem: can end up with unphysical uncertainties (e.g. $m = 0.1 \pm 0.3$kg).

10 Bayesian Estimation

Bayes' Theorem:

$$P(\theta|Data) = \frac{P(Data|\theta)P(\theta)}{P(Data)}, \tag{23}$$

where $P(\theta)$ is the *prior*, our initial guess (often *flat*, i.e. uniform), $P(\theta|Data)$ is the posterior, the estimate given the data. By the marginalisation rule:

$$P(Data) = \int P(Data|\theta)P(\theta)d\theta. \tag{24}$$

Physical constraints: make prior zero in unphysical regions.