SOME DISCRETE DISTRIBUTIONS

Name	**Parameters**	**Genesis / Usage**	$p(x) = \mathbb{P}[X = x]$ **and non-zero range**	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	**Comments**
Uniform (discrete)	$k \in \mathbb{N}$	Set of k equally likely outcomes.	$p(x) = 1/k$ $x = 1, ..., k$	$\frac{k+1}{2}$	$\frac{k^2-1}{12}$	Fair dice roll with $k = 6$.
Bernoulli trial	$\theta \in [0, 1]$	Experiment with two outcomes; typically, success = 1, fail = 0.	$p(x) = \theta^x(1-\theta)^{1-x}$ $x = 0, 1$	θ	$\theta(1-\theta)$	
Binomial	$n \in \mathbb{N}$ $\theta \in [0, 1]$	Number of successes in n i.i.d. Bernoulli trials.	$p(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ $x = 0, 1, 2, ..., n$	$n\theta$	$n\theta(1-\theta)$	Often written $\mathrm{Bin}(n, \theta)$. $\mathrm{Bin}(1, \theta) \sim \mathrm{Bernoulli}(\theta)$
Geometric	$\theta \in (0, 1]$	Number of failed i.i.d. Bernoulli trials before the first success.	$p(x) = \theta(1-\theta)^x$ $x = 0, 1, 2, \ldots$	$\frac{\theta}{1-\theta}$	$\frac{\theta^2}{(1-\theta)^2}$	Alternative parametrisations: swap θ and $1-\theta$, or $X' = X + 1$ to include the final trial.
Negative Binomial	$k \in \mathbb{N}$ $\theta \in (0, 1]$	Number of failed i.i.d. Bernoulli trials before the k^{th} success.	$p(x) =$ $\binom{x+k-1}{x}\theta^k(1-\theta)^x$ $x = 0, 1, 2, \ldots$	$\frac{k(1-\theta)}{\theta}$	$\frac{k(1-\theta)}{\theta^2}$	Many alternative parametrisations. $\mathrm{NegBin}(1, \theta) \sim \mathrm{Geometric}(\theta)$.
Hypergeometric	$N \in \mathbb{N}$ $k \in \{0, \ldots, N\}$ $n \in \{0, \ldots, n\}$	Number of special objects in a random sample of n objects, from a population of N objects with k special objects.	$p(x) = \binom{k}{x}\binom{N-k}{n-x}/\binom{N}{n}$ $x = 0, ..., n$	$\frac{nk}{N}$	$n\frac{N-n}{N-1}\frac{k}{N}\times$ $(1-\frac{k}{N})$	
Poisson	$\lambda \in (0, \infty)$	Counting events occurring uniformly at random within space or time.	$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, 2, ...$	λ	λ	

SOME CONTINUOUS DISTRIBUTIONS

Name	**Parameters**	**Genesis / Usage**	$f(x) =$ **p.d.f. and non-zero range**	$\mathbb{E}[X]$	$\mathrm{Var}(X)$	**Comments**
Uniform (continuous)	$\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$	The uniform distribution for a continuous interval.	$f(x) = \frac{1}{\beta-\alpha}$ $x \in (\alpha, \beta)$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$	
Normal	$\mu \in \mathbb{R}$ $\sigma \in (0,\infty)$	Empirically and theoretically (via CLT) a good model in many situations.	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $x \in \mathbb{R}$	μ	σ^2	Often written $\mathrm{N}(\mu, \sigma^2)$. Alternative parameter: $\tau = \frac{1}{\sigma^2}$. $a\,\mathrm{N}(\mu,\sigma^2) + b \sim N(a\mu + b, a^2\sigma^2)$
Exponential	$\lambda \in (0,\infty)$	Inter-arrival times of random events.	$f(x) = \lambda e^{-\lambda x}$ $x \in (0,\infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Often written $\mathrm{Exp}(\lambda)$. Alternative parameter: $\theta = \frac{1}{\lambda}$.
Gamma	$\alpha \in (0,\infty)$ $\beta \in (0,\infty)$	Lifetimes of ageing items, multi-inter-arrival times.	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ $x \in (0,\infty)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	Often written $\Gamma(\alpha, \beta)$. Alternative parameter: $\theta = \frac{1}{\beta}$. $\mathrm{Gamma}(1, \lambda) \sim \mathrm{Exp}(\lambda)$
Beta	$\alpha \in (0,\infty)$ $\beta \in (0,\infty)$	Quantities constrained to be within intervals.	$f(x) = \frac{1}{\mathcal{B}(a,b)} x^{\alpha-1}(1-x)^{\beta-1}$ $x \in [0,1]$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$	$\mathrm{Beta}(1,1) \sim \mathrm{Uniform}(0,1)$
Cauchy	$a \in \mathbb{R}$ $b \in (0,\infty)$	Heavy tailed, pathological examples.	$f(x) = \frac{1}{\pi b} \frac{b^2}{(x-a)^2 + b^2}$ $x \in \mathbb{R}$	undefined	undefined	
Pareto	$\alpha \in (0,\infty)$ $\beta \in (0,\infty)$	Heavy tailed quantities.	$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}$ $x \in (\beta, \infty)$	$\frac{\alpha\beta}{\alpha-1}$ if $\alpha > 1$	$\frac{\alpha^2\beta}{(\alpha-1)^2(\alpha-2)}$ if $\alpha > 2$	Sometimes written $\mathrm{Pareto}(\beta, \alpha)$. $\log\left(\frac{\mathrm{Pareto}(\alpha,\beta)}{\beta}\right) \sim \mathrm{Exp}(\alpha)$
Weibull	$\lambda \in (0,\infty)$ $k \in (0,\infty)$	Lifetimes, extreme values.	$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$ $x \in (0,\infty)$	$\lambda\Gamma(1 + 1/k)$	$\lambda^2 [\Gamma(1 + 2/k)$ $+ \Gamma(1 + 1/k)^2]$	$\left(\frac{\mathrm{Weibull}(\lambda,k)}{\lambda}\right)^k \sim \mathrm{Exp}(1)$
Log-Normal	$\mu \in \mathbb{R}$ $\sigma \in (0,\infty)$	Quantities related to exponential growth.	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{\sqrt{2}\sigma}\right)$ $x \in (0,\infty)$	$e^{\mu + \frac{1}{2}\sigma^2}$	$(e^{\sigma^2} - 1)$ $\times e^{2\mu+\sigma^2}$	Often written $\mathrm{LogN}(\mu, \sigma^2)$. $\log(\mathrm{LogN}(\mu,\sigma^2)) \sim N(\mu, \sigma^2)$
Chi-squared	$n \in \mathbb{N}$	Statistical testing.	$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$ $x \in (0,\infty)$	n	$2n$	Often written χ^2_n. $X^2_n \sim \mathrm{Gamma}(n/2, 1/2)$ $X_i \sim N(0,1)$ i.i.d. $\Rightarrow \sum_1^n X_i^2 \sim \chi^2_n$
Student t	$n \in \mathbb{N}$	Statistical testing.	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$ $x \in \mathbb{R}$	0 if $n > 1$	$\frac{n}{n-2}$ if $n > 2$	Often written t_n. Can allow $n \in (0,\infty)$. $t_1 \equiv \mathrm{Cauchy}(0,1)$
Inverse Gamma	$\alpha \in (0,\infty)$ $\beta \in (0,\infty)$	Quantities related to the Gamma distribution.	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\beta/x)$ $x \in (0,\infty)$	$\frac{\beta}{\alpha-1}$ if $\alpha > 1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ if $\alpha > 2$	Often written $\mathrm{IGamma}(\alpha, \beta)$. $\mathrm{IGamma}(\alpha,\beta) \sim \frac{1}{\mathrm{Gamma}(\alpha,\beta)}$

SOME CONJUGATE PAIRS

Model family	**Prior family**	**Data**	**Posterior parameters**
$\text{Bernoulli}(\theta)^{\otimes n}$	$\theta \sim \text{Beta}(a,b)$	$x \in \{0,1\}^n$	$a^* = a + \sum_1^n x_i$ $b^* = b + n - \sum_1^n x_i$
$\text{Bin}(m_1,\theta) \otimes \ldots \otimes \text{Bin}(m_n,\theta)$ with $m_1,\ldots m_n \in \mathbb{N}$ fixed.	$\theta \sim \text{Beta}(a,b)$	$x \in \{0,1,\ldots,\}^n$ where $x_i \in \{0,\ldots,m_i\}$	$a^* = a + \sum_1^n x_i$ $b^* = b + \sum_1^n m_i - \sum_1^n x_i$
$\text{Geometric}(\theta)^{\otimes n}$	$\theta \sim \text{Beta}(a,b)$	$x \in \{0,1,\ldots,\}^n$	$a^* = a + n$ $b^* = b + \sum_1^n x_i$
$\text{Poisson}(\theta)^{\otimes n}$	$\theta \sim \text{Gamma}(a,b)$	$x \in \{0,1,\ldots,\}^n$	$a^* = a + \sum_1^n x_i$ $b^* = b + n$
$\text{Exp}(\lambda)^{\otimes n}$	$\lambda \sim \text{Gamma}(a,b)$	$x \in (0,\infty)^n$	$a^* = a + n$ $b^* = b + \sum_1^n x_i$
$\text{Weibull}(\theta,\beta)^{\otimes n}$ with $\beta \in (0,\infty)$ fixed.	$\theta \sim \text{IGamma}(a,b)$	$x \in (0,\infty)^n$	$a^* = a + n$ $b^* = b + \sum_1^n x_i^\beta$
$\text{N}(\theta,\sigma^2)^{\otimes n}$ with $\sigma \in (0,\infty)$ fixed.	$\theta \sim \text{N}(u,s^2)$	$x \in \mathbb{R}^n$	$u^* = \left(\frac{1}{\sigma^2}\sum_1^n x_i + \frac{u}{s^2}\right) / \left(\frac{n}{\sigma^2} + \frac{1}{s^2}\right)$ $(s^*)^2 = 1 / \left(\frac{n}{\sigma^2} + \frac{1}{s^2}\right)$
$\text{N}(\theta,\frac{1}{\tau})^{\otimes n}$ with $\tau \in (0,\infty)$ fixed.	$\theta \sim \text{N}(u,\frac{1}{t})$	$x \in \mathbb{R}^n$	$u^* = (\tau\sum_1^n x_i + ut) / (\tau n + t)$ $\frac{1}{t^*} = 1 / (\tau n + t)$
$\text{N}(\mu,\frac{1}{\tau})^{\otimes n}$ with $\mu \in \mathbb{R}$ fixed.	$\tau \sim \text{Gamma}(a,b)$	$x \in \mathbb{R}^n$	$a^* = a + \frac{n}{2}$ $b^* = b + \frac{1}{2}\sum_1^n (x_i - \mu)^2$
$\text{N}(\mu,\frac{1}{\tau})^{\otimes n}$	$(\mu,\tau) \sim \text{NGamma}(m,p,a,b)$	$x \in \mathbb{R}^n$	$m^* = \frac{n\bar{x}+mp}{n+p}$ $p^* = n + p$ $a^* = a + \frac{n}{2}$ $b^* = b + \frac{n}{2}\left(s^2 + \frac{p}{n+p}(\bar{x}-m)^2\right)$ where $\bar{x} = \frac{1}{n}\sum_1^n x_i$ and $s^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$

See the sheet on conditional probability for the Normal-Gamma distribution.
For all other distributions, see the reference sheets of discrete and continuous distributions.

CONDITIONAL PROBABILITY AND RELATED FORMULAE

We say that a random variable X is **discrete** if there exists a countable set $A \subseteq \mathbb{R}^d$ such that $\mathbb{P}[X \in A] = 1$. In this case the function $p_X(x) = \mathbb{P}[X = x]$, defined for $x \in \mathbb{R}^d$, is known as the **probability mass function** of X. The **range** of X is the set $R_X = \{x \in \mathbb{R}^d\,;\, \mathbb{P}[X = x] > 0\}$.

We say that a random variable X is **continuous** if there exists a function $f_X : \mathbb{R}^d \to [0, \infty)$ such that $\mathbb{P}[X \in A] = \int_A f_X(x)\,dx$ for all $A \subseteq \mathbb{R}^d$. In this case f_X is known as the **probability density function** of X. The **range** of X is the set $R_X = \{x \in \mathbb{R}^d\,;\, f_X(x) > 0\}$.

If X and Y are discrete, and $p_X \propto p_Y$, then $X \stackrel{d}{=} Y$.
If X and Y are continuous, and $f_X \propto f_Y$, then $X \stackrel{d}{=} Y$.

If X is a random variable and $\mathbb{P}[X \in A] > 0$ then the **conditional distribution** of $X|_{\{X\in A\}}$ satisfies $\mathbb{P}[X|_{\{X\in A\}} \in A] = 1$ and

$$\mathbb{P}[X|_{\{X\in A\}} \in B] = \frac{\mathbb{P}[X \in B]}{\mathbb{P}[X \in A]}$$

for all $B \subseteq A$.

If X and Y are random variables, with $A \subseteq R_X$, $B \subseteq R_Y$ and $\mathbb{P}[X \in A] > 0$, then

$$\mathbb{P}[Y|_{\{X\in A\}} \in B] = \frac{\mathbb{P}[X \in A, Y \in B]}{\mathbb{P}[X \in A]}.$$

If (Y, Z) and random variables and $\mathbb{P}[Y = y] = 0$ then it is sometimes possible to define the conditional distribution of $Z|_{\{Y=y\}}$ via taking the limit $\mathbb{P}\left[Z|_{\{|Y-y|\leq\epsilon\}} \in A\right] \to \mathbb{P}[Z|_{\{Y=y\}} \in A]$ as $\epsilon \to 0$.

Let (Y, Z) be a pair of continuous random variables. If the conditional distribution of $Z|_{\{Y=y\}}$ exists then it is given by

$$f_{Z|_{\{Y=y\}}}(z) = \frac{f_{Y,Z}(y, z)}{f_Y(y)}.$$

For a discrete or continuous random variable X, the **likelihood function** of X is

$$L_X(x) = \begin{cases} \mathbb{P}[X = x] & \text{if } X \text{ is discrete,} \\ f_X(X) & \text{if } X \text{ is continuous.} \end{cases}$$

The general formula for **completing the square** as a function of $\theta \in \mathbb{R}$ is $A\theta^2 - 2\theta B + C = A\left(\theta - \frac{B}{A}\right)^2 + C - \frac{B^2}{A}$

The **sample-mean-variance** identity states $\sum_1^n (x_i - \mu)^2 = ns^2 + n(\bar{x} - \mu)^2$ where $\bar{x} = \frac{1}{n}\sum_1^n x_i$ and $s^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$.

The **Beta and Gamma functions** are given by

$$\mathcal{B}(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}\,dx, \qquad \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt.$$

They are related by $\mathcal{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. For $n \in \mathbb{N}$, $(n - 1)! = \Gamma(n)$.

The **Normal-Gamma distribution** has p.d.f. given by

$$\begin{aligned} f_{\mathrm{NGamma}(m,p,a,b)}(\mu, \tau) &= f_{\mathrm{N}(m,\frac{1}{p\tau})}(\mu)\, f_{\mathrm{Gamma}(a,b)}(\tau) \\ &\propto \tau^{a-\frac{1}{2}} \exp\left(-\frac{p\tau}{2}(\mu - m)^2 - b\tau\right). \end{aligned}$$

for $\mu \in \mathbb{R}$ and $\tau > 0$, and zero otherwise. The parameters are $m \in \mathbb{R}$, $p \in (0, \infty)$, $a \in (0, \infty)$ and $b \in (0, \infty)$. If $(U, T) \sim \mathrm{NGamma}(m, p, a, b)$ then $T \sim \mathrm{Gamma}(a, b)$ and $U|_{\{T=\tau\}} \sim \mathrm{N}(m, \frac{1}{p\lambda})$.

BAYESIAN MODELS AND RELATED FORMULAE

The **Bayesian model** associated to the model family $(M_\theta)_{\theta\in\Pi}$ and prior p.d.f. $f_\Theta(\theta)$ is the random variable $(X,\Theta)\in\mathbb{R}^n\times\mathbb{R}^d$ with distribution given by

$$\mathbb{P}[X\in B,\Theta\in A]=\int_A \mathbb{P}[M_\theta\in B]f_\Theta(\theta)\,d\theta.$$

The model family satisfies $X|_{\{\Theta=\theta\}}\stackrel{\mathrm{d}}{=}M_\theta$.

The distribution of X is known as the **sampling distribution**, given by

$$\begin{aligned} \mathbb{P}[X=x] &= \int_{\mathbb{R}^d}\mathbb{P}[M_\theta=x]f_\Theta(\theta)\,d\theta \qquad &&\text{if } (M_\theta) \text{ is a discrete family,}\\ f_X(x) &= \int_{\mathbb{R}^d} f_{M_\theta}(x)f_\Theta(\theta)\,d\theta. &&\text{if } (M_\theta) \text{ is a continuous family.}\end{aligned}\qquad(\star)$$

The distribution of $\Theta|_{\{X=x\}}$ is known as the **posterior distribution** given the data x. **Bayes rule** states that

$$f_{\Theta|_{\{X=x\}}}(\theta)=\frac{1}{Z}L_{M_\theta}(x)f_\Theta(\theta)$$

where L_{M_θ} is the likelihood function of M_θ; the p.d.f. in the absolutely continuous case and the p.m.f. in the discrete case. The normalizing constant Z is given by $Z=\int_\Pi L_{M_\theta}(x)f_\Theta(\theta)\,d\theta$, which is equal to $\mathbb{P}[X=x]$ in the discrete case and equal to $f_X(x)$ is the continuous case.

The **predictive distribution** is given by replacing f_Θ in $(\star)$ with $f_{\Theta|_{\{X=x\}}}$.

If θ is a real valued parameter and $X\sim M_\theta$, the **reference prior** Θ associated to the model family (M_θ) has density function given by

$$f_\Theta(\theta)\propto\mathbb{E}\left[\left(\frac{d}{d\theta}\log(L_{M_\theta}(X))\right)^2\right]^{1/2}\propto\mathbb{E}\left[-\frac{d^2}{d\theta^2}\log(L_{M_\theta}(X))\right]^{1/2}.$$

Consider a Bayesian model with unknown parameter θ and data x. Let H_0 be the hypothesis that $\theta\in\Pi_0$, and H_1 be the hypothesis that $\theta\in\Pi_1$, where Π_0 and Π_1 partition the parameter space Π. The **prior and posterior odds ratios** of H_0 against H_1 are

$$\frac{\mathbb{P}[\Theta\in\Pi_0]}{\mathbb{P}[\Theta\in\Pi_1]}\qquad\text{and}\qquad\frac{\mathbb{P}[\Theta|_{\{X=x\}}\in\Pi_0]}{\mathbb{P}[\Theta|_{\{X=x\}}\in\Pi_1]}.$$

The **Bayes factor** is $B=\frac{\text{posterior odds}}{\text{prior odds}}$. The following table provides a rough guide to interpreting the Bayes factor.

Bayes factor	Interpretation: evidence in favour of H_0 over H_1
1 to 3.2	Indecisive / not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
above 100	Decisive

A **high posterior density region** is a subset $\Pi_0\subseteq\Pi$ that is chosen to minimize the size of Π_0 and maximize $\mathbb{P}[\Theta|_{\{X=x\}}\in\Pi_0]$.

If $\Theta|_{\{X=x\}}$ has a distribution with a single peak then it is common to choose an **equally tailed** HPD region of the form $\Pi_0=[a,b]$ where

$$\mathbb{P}\left[\Theta|_{\{X=x\}}<a\right]=\mathbb{P}\left[\Theta|_{\{X=x\}}>b\right]=\frac{1-p}{2}$$

and some value is picked for $p\in(0,1)$.

If $Z\sim\mathrm{N}(0,1)$ then $\mathbb{P}[Z\geq 1.645]\approx 0.95$, $\mathbb{P}[Z\geq 1.96]\approx 0.975$ and $\mathbb{P}[Z\geq 2.58]\approx 0.995$.

SOME USEFUL ALGORITHMS

The **Metropolis-Hastings** algorithm for simulating (approximate) samples from the distribution of Y is as follows. The key ingredient of the algorithm is a joint distribution (Y, Q), where $Q|_{\{Y=y\}}$ and $Y|_{\{Q=y\}}$ are both well defined for all $y \in R_Y$, both with the same range as Y.

Let y_0 be a point within R_Y. Then, given y_m we define y_{m+1} as follows.

1. Generate a *proposal point* $\tilde{y}$ from the distribution of $Q|_{\{Y=y_m\}}$.
2. Calculate the value of $\alpha = \min\left\{1, \frac{f_{Y|_{\{Q=\tilde{y}\}}}(y_m) f_Y(\tilde{y})}{f_{Q|_{\{Y=y_m\}}}(\tilde{y}) f_Y(y_m)}\right\}$.
3. Then, set $y_{m+1} = \begin{cases} \tilde{y} & \text{with probability } \alpha, \\ y_m & \text{with probability } 1-\alpha. \end{cases}$

For sufficiently large m, the distribution of y_m is approximately that of Y.

The distribution $Q|_{\{Y=y\}}$ is called the *proposal* distribution, based on its role in steps 1 and 2. The two cases in step 3 are usually referred to as *acceptance* (when $y_{m+1} = \tilde{y}$) and *rejection* (when $y_{m+1} = y_m$).

The **Metropolis** algorithm is the special case

$$f_{Q|_{\{Y=y\}}}(\tilde{y}) = f_{Y|_{\{Q=\tilde{y}\}}}(y), \tag{†}$$

in which case step 2 simplifies to $\alpha = \min\left\{1, \frac{f_Y(\tilde{y})}{f_Y(y_m)}\right\}$.

The **random walk Metropolis** algorithm is the choice $Q = Y + Z$, where Z is independent of Y and Q and satisfies $f_Z(z) = f_Z(-z)$ for all $z \in R_Z$. In this case

$$Q|_{\{Y=y\}} \stackrel{\mathrm{d}}{=} y + Z \qquad \text{and} \qquad Y|_{\{Q=\tilde{y}\}} \stackrel{\mathrm{d}}{=} \tilde{y} + Z,$$

which implies (†). A common choice is $Z \sim \mathrm{N}(0, \sigma^2)$.

The **random walk MCMC algorithm** is obtained by applying the random walk Metroplis algorithm to find the posterior distribution of a Bayesian model. The algorithm is as follows. We start with a (discrete or continuous) Bayesian model (X, Θ), where the parameter space is $\Pi = \mathbb{R}^d$. We want to obtain samples of $\Theta|_{\{X=x\}}$ and we know the p.d.f. $f_{\Theta|_{\{X=x\}}}$.

Choose an initial point $y_0 \in \Pi$. Choose a continuous distribution for Z satisfying $f_Z(z) = f_Z(-z)$ for all $z \in \mathbb{R}$. A common choice is $Z \sim N(0, \sigma^2)$.

Then, given y_m, we define y_{m+1} as follows.

1. Sample z from Z and set $\tilde{y} = y_m + z$.
2. Calculate $\alpha = \min\left(1, \frac{f_{\Theta|_{\{X=x\}}}(\tilde{y})}{f_{\Theta|_{\{X=x\}}}(y_m)}\right)$.
3. Then, set $y_{m+1} = \begin{cases} \tilde{y} & \text{with probability } \alpha, \\ y_m & \text{with probability } 1-\alpha. \end{cases}$

The **Gibbs sampler** for $\theta = (\theta_1, \ldots, \theta_d)$ is as follows. We first choose an initial point $y_0 = (\theta_1^{(0)}, \ldots, \theta_d^{(0)}) \in \Pi$. Then, for each $i = 1, \ldots, d$, sample $\tilde{y}$ from $\Theta_{-i}|_{\{X=x\}}$ and set

$$y_{m+1} = (\theta_1^{(m)}, \ldots, \theta_{i-1}^{(m)}, \tilde{y}, \theta_{i+1}^{(m)}, \ldots, \theta_d^{(m)}).$$

Note that we increment the value of m each time that we increment i. When reach $i = d$, return to $i = 1$ and repeat. For sufficiently large m, the distribution of y_m is approximately that of $\Theta|_{\{X=x\}}$.

The distributions of $\Theta_i|_{\{\Theta_{-i}=\theta_{-i}, X=x\}}$, for $i = 1, \ldots, d$, are known as the **full conditional distributions** of Θ. They satisfy

$$f_{\Theta_i|_{\{\Theta_{-i}=\theta_{-i}, X=x\}}}(\theta_i) \propto f_{\Theta|_{\{X=x\}}}(\theta)$$

Here $\propto$ treats θ_{-i} and x as constants, and the only variable is θ_i.