# Common Distributions

The notation $z \sim \mathrm{P}(\theta)$ means that the random variable $z$ is sampled from (or distributed according to) the distribution P, which is parameterized by $\theta$. When we write $\mathrm{P}(z;\theta)$ we refer to the density (assuming it exists) of P evaluated at $z$. Here, we provide a summary of common distributions and their parametric densities or mass functions.

**Bernoulli** For a binary random variable $x \in \{0,1\}$ with $\rho \in [0,1]$,

$$\mathrm{Bern}(x;\rho) = \rho^x (1-\rho)^{1-x}.$$

Sometimes we parameterize the Bernoulli distribution in terms of its *logit*, $\sigma^{-1}(\rho) = \log \frac{\rho}{1-\rho}$, where $\sigma(u) = (1+e^{-u})^{-1}$ is the *logistic/sigmoid* function.

**Beta** For a continuous random variable $\rho \in [0,1]$ with $a > 0$ and $b > 0$,

$$\mathrm{Beta}(\rho; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho^{a-1}(1-\rho)^{b-1}.$$

Note that for $\alpha < 1$ and/or $\beta < 1$, the beta density diverges and is **unbounded**.

**Binomial** For an integer-valued random variable $x \in \{1,\ldots,N\}$ with $N \in \mathbb{N}$ and $\rho \in [0,1]$,

$$\mathrm{Bin}(x; N, \rho) = \binom{N}{x} \rho^x (1-\rho)^{N-x}.$$

**Categorical** For an integer-valued random variable $x \in \{1,\ldots,K\}$ with $K \in \mathbb{N}$ and $\boldsymbol{\pi} \in \Delta_K$,

$$\mathrm{Cat}(x; \boldsymbol{\pi}) = \pi_x,$$

where $\pi_x$ denotes the $x$-the coordinate of the vector $\boldsymbol{\pi}$.

*Note: sometimes the categorical distribution is used as a distribution on one-hot vectors* $\boldsymbol{x} \in \{0,1\}^K$; $\sum_{k=1}^K x_k = 1$, *in which case the pmf is* $\mathrm{Cat}(\boldsymbol{x}; \boldsymbol{\pi}) = \prod_{k=1}^K \pi_k^{x_k}$. *It will be clear from context if* $x$ *is an integer or a one-hot vector.*

**Dirichlet** For a probability vector $\boldsymbol{\pi} \in [0,1]^K$ such that $\pi_k \geq 0$ and $\sum_k \pi_k = 1$, and parameter $\boldsymbol{\alpha} \in \mathbb{R}_+^K$,

$$\mathrm{Dir}(\boldsymbol{\pi}; \boldsymbol{\alpha}) = \frac{\Gamma\big(\sum_{k=1}^K \alpha_k\big)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \pi_k^{\alpha_k - 1}.$$

**Exponential** For a random variable $x \in \mathbb{R}_+$ with rate $\lambda \in \mathbb{R}_+$,

$$\text{Exp}(x; \lambda) = \lambda e^{-\lambda x}\, \mathbb{I}[x \geq 0].$$

The cumulative distribution function for the exponential is $\Pr(x < a; \lambda) = 1 - e^{-\lambda a}$. You can sample an exponential distribution by first sampling $u \sim \text{Unif}([0,1])$, then setting $x = -\frac{\log u}{\lambda}$.

**Gamma** For a nonnegative random variable $\lambda \in \mathbb{R}_+$ with shape parameter $a > 0$ and rate parameter $b > 0$,

$$\text{Ga}(\lambda; a, b) = \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}.$$

The gamma distribution may also be parameterized in terms of a scale parameter, $\theta = b^{-1}$, but we do not use that parameterization in this exam.

**Multinomial** For a vector of discrete counts $\boldsymbol{x} \in \mathbb{N}^K$ with $\sum_k x_k = N$ and a probability vector $\boldsymbol{\pi} \in [0,1]^K$,

$$\text{Mult}(\boldsymbol{x}; N, \boldsymbol{\pi}) = \binom{N}{x_1, x_2, \ldots x_K} \prod_{k=1}^{K} \pi_k^{x_k},$$

where

$$\binom{N}{x_1, x_2, \ldots x_K} = \frac{N!}{x_1! \ldots x_K!}.$$

**Multivariate Normal** For a random variable $\boldsymbol{x} \in \mathbb{R}^D$ with mean $\boldsymbol{\mu} \in \mathbb{R}^D$ and positive semidefinite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$,

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-D/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}.$$

**Negative Binomial** For an integer-valued random variable $x \in \mathbb{N}$ with shape parameters $\nu \in \mathbb{R}_+$ and probability $\rho \in [0,1]$,

$$\text{NB}(x; \nu, \rho) = \binom{x + \nu - 1}{x} \rho^x (1-\rho)^\nu.$$

**Poisson** For an integer random variable $x \in \mathbb{N}$ and a nonnegative rate parameters $\lambda \in \mathbb{R}_+$,

$$\text{Po}(x; \lambda) = \frac{1}{x!} \lambda^x e^{-\lambda}.$$

The mean and variance of the Poisson distribution are both equal to $\lambda$.

**Truncated Normal** For a random variable $x \in [a, b)$ and parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_+$,

$$\text{TruncNorm}(x; \mu, \sigma^2, [a,b)) = \frac{\mathcal{N}(x; \mu, \sigma^2)}{\int_a^b \mathcal{N}(y; \mu, \sigma^2)\, \mathrm{d}y} \mathbb{I}[a \leq x < b].$$

**Uniform** For a continuous random variable $x \in [a, b)$,

$$\text{Unif}(x; [a,b)) = \frac{1}{b-a} \mathbb{I}[a \leq x < b].$$