Advanced probability distributions

Measurements

Laplace



- μ location: mean and median
- b scale: average distance of y from μ

All real numbers. Sharper peak and heavier tails than Normal. Regularisation prior (Bayesian LASSO).

Cauchy



- μ location: median
- γ scale; half of the values lie within $\mu \pm \gamma$

All real numbers. Very heavy tails, no mean. Student-t with $\nu = 1$. Often truncated at 0 $\rightarrow$ variance prior.

Log-normal

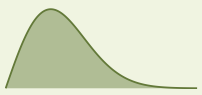


- μ mean of $\log y$; median is e^{μ}
- σ standard deviation of $\log y$

Positive reals. Skewed positive data with multiplicative effects, e.g. reaction time, income.

Waiting times

Weibull



- k shape: $k < 1 \rightarrow$ falling risk, $k > 1 \rightarrow$ rising risk
- λ scale

Positive reals. Time to event with risk changing over time. Survival models. Exponential at $k = 1$.

Outcomes & choices

Ordered logistic



- η mean on the latent scale
- c $K - 1$ ordered cutpoints between categories

Integers 1 to K. Ordered categories, e.g. Likert ratings. Categories need not be equidistant.

Scales & variances

InvGamma



- α shape
- β scale; mean is $\beta/(\alpha - 1)$ if $\alpha > 1$

Positive reals. Prior for variances. Disallows values close to zero $\rightarrow$ careful with hierarchical models!

Correlated quantities

MvNormal



- μ mean vector, one entry for each of K dimensions
- Σ $K \times K$ covariance matrix

Vectors of K reals. Correlated outcomes and correlated parameters. Σ may involve LKJ prior.

LKJ

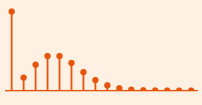


- K number of rows and columns
- η $\eta = 1$ is uniform; larger $\eta \rightarrow$ lower correlations

$K \times K$ correlation matrices Ω; prior for correlation in MvNormal. Often better to sample Cholesky factor.

Extensions

Mixtures



- p simplex of length K; components probabilities
- d vector of K component distributions

K processes generate data. Example: mix Poisson + excess zeros (Dirac) $\rightarrow$ zero-inflated Poisson.

Truncation



- l lower bound
- u upper bound

Limits support to $[l, u]$. Good for variance priors. Careful otherwise: this introduces strong assumptions!

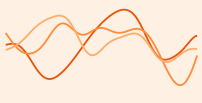
Censoring



- l lower limit; smaller values are recorded as l
- u upper limit; larger values are recorded as u

Limits which parts of the distribution are observed and can be learned about. Closely related to missing data.

Gaussian process



- m mean function, often 0
- k covariance kernel: correlation of $f(x)$ and $f(x')$

Prior over functions. Nonparametric. Time and space. Scales poorly at $\mathcal{O}(N^3) \rightarrow$ consider approximations.