II.3.2 Use of Parametric Distributions

1. Note: As noted above, we are interested in probability in order to describe data distributions or to describe the optimality and precision of statistical inference. We thus can highlight some special probability distributions according to their general use in these two roles.

2. Note: In describing data distributions, it is useful to consider scientific mechanisms that might lead to distributions that are well approximated by particular distribution families.

3. Note: "Scientifically motivated" discrete distributions include

   a.) Bernoulli (discrete binary variables)

   b.) Binomial (counts events in n independent Bernoulli observations)

   c.) Geometric (number of observations prior to and including the first event in independent Bernoulli observations)

   d.) Negative Binomial (number of observations prior to and including the kth event in independent Bernoulli observations)

   e.) Multinomial (counts observations falling in several unordered categories)

   f.) Poisson (count of events observed in a process occurring at a constant rate over space and time)

   g.) Discrete Uniform (equally likely outcomes in an ordered discrete sample space)

4. Note: "Scientifically motivated" continuous distributions include

   a.) Normal (outcomes that reflect the sum of multiple inputs)

   b.) Lognormal (outcomes that reflect the product of multiple inputs)

   c.) Exponential (times to events when hazards are approximately constant)

   d.) Gamma (times to failure when components with constant hazard are in parallel)

   e.) Weibull (times to failure when components with constant hazard are in series or hazards are approximately log linear in log time)

5. Note: "Statistically motivated" distributions used in deriving candidate analysis models (estimating equations)

   a.) Exponential family (including normal, binomial, Poisson, exponential)

   b.) Accelerated failure time models (including exponential, Weibull, gamma, log gamma, lognormal)

   c.) Location-scale semi parametric families (including normal)

   d.) Proportional hazards families (including exponential, Weibull)

6. Note: "Statistically motivated" sampling distributions include

   a.) Normal (statistics most often involve sums so CLT is relevant)

   b.) Chi square (quadratic forms calculated from multivariate statistics)
   c.) Hypergeometric (used in Fisher's exact test– a conditional test sometimes used in 2x2 tables)
   d.) Gamma (sums of log transformed p values)
   e.) t and F (sampling distributions used as small sample approximations for commonly used statistics)
   f.) Uniform (for distribution of p values under a null hypothesis)

7. Note: "Statistically motivated" conjugate distributions include
   a.) Normal (conjugate for normally distributed data)
   b.) Gamma (conjugate for exponential and gamma distributed data)
   c.) Beta (conjugate for binomial)
   d.) Dirichlet (conjugate for multinomial)

8. Note: "Statistically motivated" distributions when exploring robustness of statistical methods either analytically or via simulation
   a.) As a function of skewness: Bernoulli, lognormal, gamma, mixtures of normals
   b.) As a function of kurtosis: t

9. Note: "Educationally motivated" distributions when learning statistical inference, including variations of all the above, plus the shifted gamma, the double exponential.