Discrete Distribution Families

Binomial(n, p)*binom

successes in n Bernoulli trials.
PMF: $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$
Mean: np, Var: $np(1-p)$

Geometric(p)*geom

Failures before 1st success.
PMF: $P(X=x) = (1-p)^x p$, $x \geq 0$
Mean: $(1-p)/p$, Var: $(1-p)/p^2$

NegBin(k, p)*nbinom

Failures before k-th success.
PMF: $P(X=x) = \binom{x+k-1}{x} p^k (1-p)^x$
Mean: $k(1-p)/p$, Var: $k(1-p)/p^2$

Poisson(λ)*pois

Counts in interval, rate λ.
PMF: $P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$
Mean: λ, Var: λ

Continuous Distribution Families

Uniform(a, b) - *unif

All values equally likely on $[a, b]$.
PDF: $f(x) = \frac{1}{b-a}$, $a \leq x \leq b$
Mean: $(a+b)/2$, Var: $(b-a)^2/12$
For a uniform distribution between a and b,

$$P(X > x) = \frac{b-x}{b-a}, \quad a \leq x \leq b$$

Normal(μ, σ^2) - *norm

Bell-shaped, $-\infty < \mu < \infty, \sigma^2 > 0$
PDF: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}$
Mean: μ, Var: σ^2
Z score: $\mu = x - Z\sigma$

Lognormal(μ, σ^2) - *lnorm

If $\ln X \sim N(\mu, \sigma^2)$.
PDF: $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x-\mu)^2/(2\sigma^2)}$
Mean: $e^{\mu+\sigma^2/2}$, Var: $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$
Skewness: $(e^{\sigma^2}+2)\sqrt{e^{\sigma^2}-1}$

Exponential(λ) - *exp

Time between Poisson events, Positive RV, wait time, memoryless
(λ) is average rate (β) is mean wait time
PDF: $f(x) = \lambda e^{-\lambda x}, x \geq 0$
Mean: $1/\lambda$, Var: $1/\lambda^2$
Skewness: 2

Beta(α, β) - *beta

On $[0, 1]$, Bayesian priors. Uniform distribution special case
(Γ) function is a generalization of factorial function to non-integer numbers.
PDF: $f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Mean: $\alpha/(\alpha+\beta)$, Var: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Skewness: $\frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$

Weibull(k, λ) - *weibul

Lifetimes, survival, reliability.
Exponential family (when k = 1), longer you wait
Survival Analysis
PDF: $f(x) = \frac{k}{\lambda}(x/\lambda)^{k-1} e^{-(x/\lambda)^k}$
Mean: $\lambda\Gamma(1+1/k)$, Var: $\lambda^2[\Gamma(1+2/k) - \Gamma(1+1/k)^2]$
Skewness: $\frac{\Gamma(1+3/k)\lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$

Continuous RV (numerical integration)

$f_X(x)$ can take on values larger than 1
$f_X(x)$ cannot give probability, area under the graph does
the density of $f_X(x)$ cannot take negative values even if total area is 1 because probability cannot be negative
the support of $f_X(x)$ must be finite

```
f <- function(x) 2*x
EV <- integrate(function(x) x*f(x), 0,1)$value
```

Define the PDF function

```
f <- function(x)  2*x
prob <- integrate(f, lower = 0.5, upper =
0.75)$value # Integrate over [0.5, 0.75]
```

Conditional Distributions

Conditional dist is just a segment of marginal dist, then re-normalized to have an area under the curve equal to 1 If X and Y are independent, in continouse case,

$$f_{Y|X}(y) = f_Y(y)$$

This means conditional PDF of Y and X is marginal PDF of Y.

Random Sample

It is independent and identically distributed (iid). Each pair of observations are independent, and each observation comes from the same distribution.

MLE

Great way to find estimators. Applied on multi and univariate. Relies on random sample of n observations. Mean, is a estimator for univariate, multivariate, linear regression

Steps for MLE

1.(discrete or contin) distribution
2.Find parameters of a theoretical distribution (eg λ in a Poisson distribution)
3.Choose distribution)
4.Play with the parameters for that family of distributions to find the most likely parameters for the corresponding parametric estimates)
5.To get estimates, we use the likelihood function of our observed random sample.)

R: Generating Sample and Likelihood

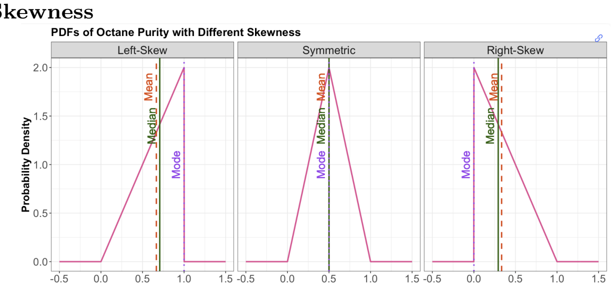
```
#Generate 30 random samples from Exponential(beta=2
set.seed(123)
sample_n30 <- tibble(values = rexp(30, rate = 1/20
#Compute likelihood and log-likelihood for candidat
exp_values <- tibble(
  possible_betas = seq(5, 50, 0.5),
  likelihood = map_dbl(1 / possible_betas,
   ~ prod(dexp(sample_n30$values, .))),
  log_likelihood = map_dbl(1 / possible_betas,
   ~  log(prod(dexp(sample_n30$values, .))))
)
empirical_MLE <- exp_values |>
  arrange(desc(likelihood)) |>
  slice(1)
analytical_MLE <- mean(sample_n30$values)
#We use the sample mean() function from R!
round(analytical_MLE, 2)
LL <- function(l) log(prod(dexp(sample_n30$values,
rate = 1 / l)))
optimize(LL, c(5, 50), maximum = TRUE)
```

stochastic - having some uncertain outcome.
deterministic - an outcome that will be known with 100% certainty.
Computers cannot actually generate truly random outcomes.
Instead, they use something called pseudorandom numbers.
Although this sequence is deterministic, it behaves like a random sample.
One pitfall is that neighbouring pairs are not independent of each other.

The empirical approach (using observed data), resulting in approximate values that improve as the sample size increases
(i.e., the frequentist paradigm!).

The Law of Large of Numbers states that, as we increase our sample size n, our empirical mean converges to the true mean we want to estimate.

Skewness



Left-skewed: long tail to the left, mean < median < mode

Right-skewed: long tail to the right, mean > median > mode