1 Quantiles

If the F is the cdf of X, then $F^{-1}(\alpha)$ is the value of x_α such that $P(X \leq x_\alpha) = \alpha$; this is called α quantile of F.

2 Student-t distribution

Gaussian distributions are sensitive to outliers. Therefore, Student-t distributions are proposed to solve this problem. The parameter of a Student-t distribution is γ, which is usually set as $\gamma > 2$, and a common value is $\gamma = 4$. A Student-t distribution will degenerate to a Gaussian distribution when $\gamma >> 5$.

3 Laplacian distribution encourages sparsity

4 Properties of Beta distributions

- When a and b are both less than 1, we get a bimodal distribution with spikes at 0 and 1.
- if a and b are both greater than 1, the distribution is unimodal with mean $\frac{a}{a+b}$ mode $\frac{a-1}{a+b-2}$ and var $\frac{ab}{(a+b)^2(a+b+1)}$

5 Pareto distribution can be used to model long-tail distributions

6 Correlation

If X and Y are independent, then $p(X, Y) = p(x)p(Y)$ and $cov(X, Y) = 0$. However, the converse is not true: uncorrelated does not imply independent.

7 Dirichlet Distribution

When extending the Beta Distribution to multivariate, it becomes Direchlet Distribution.

The following is an example when $K = 3$. $Dir(1, 1, 1)$ is a uniform distribution; $Dir(2, 2, 2)$ is a broad distribution centered at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $Dir(2, 2, 2)$ is a narrow distribution centered at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. If $\alpha_k < 1$ for all k, we get spikes at the corner of the simplex.

8 Transformation of random variables

$p_y(y) = p_x(x)|\frac{dx}{dy}|$

Multivariate change of variables: $p_y(y) = p_x(x)|det(\frac{\partial x}{\partial y})|$

9 KL divergence

$KL(p||q) = H(p,q) - H(p,p)$

10 Jensen's inequality

For any convex function f, $f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i)$.

11 Mutual information

How similar the joint distribution is to the factored distribution $p(X)p(Y)$

$$I(X;Y) = KL(p(X,Y)||p(X)p(Y))$$