- Continuous pdf:
$f_X(x) : \mathcal{X} \to [0, \infty)$, $f_X(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \cdot \Pr[x \leq X \leq x + \epsilon]$
- Continuous cdf:
$F_x(x) := \Pr[X \leq x]$ where $\Pr[a \leq X \leq b] = \int_a^b f_x(x)\,\mathrm{d}x$
- Uniform distribution on $\mathcal{X} = [u, v]$:
$f_X(x) = 1/(v - u)$ for $x \in \mathcal{X}$
- Normal distribution:
$f_X(x) = (2\pi)^{-1/2} \exp(-x^2/2)$
- Dirac delta property:
$\int_{-\infty}^{\infty} \mathrm{d}x\, b(x)\delta(x - a) = b(a)$
- Expectation value of $g(X)$ where $X \in \mathcal{X}$:
$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} \Pr[X = x] g(x)$, k'th moment of X= $\mathbb{E}[X^k]$
- Statistical distance of $X, Y \in \mathcal{X}$:
$\Delta(X, Y) = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mathbb{P}(x) - \mathbb{Q}(x)|$
- Covariance matrix K:
$K_{i,j} = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \cdot \mathbb{E}[X_j]$
Zero covariance ($K_{1,2} = K_{2,1} = 0$) does not imply X_1 and X_2 are independent.
- Marginal distribution for X when $(X, Y) \sim \mathbb{P}$:
$\Pr[X = x] = \sum_y \mathbb{P}(x, y)$
- Conditional probability for $(X, Y) \sim \mathbb{P}$:
$\Pr[X = x | Y = y] = \frac{\Pr[X=x, Y=y]}{\Pr[Y=y]} = \frac{\mathbb{P}(x,y)}{\mathbb{P}_2(y)}$
- (Shannon) Entropy rules:
    1. **Additivity**: inf of a set of indep. RVs must be the sum of indiv. inf. contents
    2. **Sub-additivity**: Total inf. content of two jointly distrib. RVs cannot exceed sum of seperate infs.
    3. **Expansibility**: Adding extra outcome of prob. 0 does not affect inf.
    4. **Normalization**: The distrib $(1/2, 1/2)$ has inf. of 1 bit.
    5. The distrib $(p, 1 - p)$ for $p \to 0$ has zero inf.
- Shannon entropy:
$H(X) = \sum_{x \in \mathcal{X}} p_x \log_2 \frac{1}{p_x}$
- Binary entropy function: 2 outcomes with prob. p and $1 - p$:
$h(p) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1-p}$
- Differential entropy for continuous RV $X \sim \rho$:
$h_{\text{diff}}(X) = -\int \mathrm{d}x\, \rho(x) \log \rho(x) = \mathbb{E}_x \log \frac{1}{\rho(x)}$
- Relative entropy (Kullback-Leibler distance):
$D(\mathbb{P}||\mathbb{Q}) = \sum_{x \in \mathcal{X}} \mathbb{P}(x) \log \frac{\mathbb{P}(x)}{\mathbb{Q}(x)}$
- Entropy of jointly distrib. RVs: $H(X, Y)$ or $H(XY)$:
$H(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{xy} \log \frac{1}{p_{xy}}$
- Conditional entropy:
$H(X|Y) = \mathbb{E}_y[H(X|Y = y)] = -\sum_{x \in \mathcal{X}} p_x \sum_{y \in \mathcal{Y}} p_{x|y} \log p_{x|y}$
$H(X|Y) = H(X, Y) - H(Y)$
- Mutual information:
$\mathbf{I} = H(X) - H(X|Y)$
$\mathbf{I} = H(Y) - H(Y|X)$
$\mathbf{I} = H(X, Y) - H(X|Y) - H(Y|X)$
$\mathbf{I} = H(X) + H(Y) - H(X, Y)$