

# Prob/Stats Cheatsheet

---

**Steve Young**

ABSTRACT: Everything I know about prob/stats/maybe information theory too..

---

## Contents

<b>1</b>	<b>Conventions</b>	<b>1</b>
<b>2</b>	<b>Distributions</b>	<b>1</b>
2.1	Gaussians	1
2.1.1	Basics	1
2.1.2	Differentiation moment trick	2
2.1.3	Gaussian with Linear Term	2
2.1.4	Multivariate Gaussians	3
2.2	Bernoulli	3
<b>3</b>	<b>Prob and stats</b>	<b>3</b>
3.1	The Rules of Probability	3
3.2	Bayes' Rule	4
3.3	Expectation and Variance	4
3.4	Central Limit Theorem	4
3.5	(Weak) Law of Large Numbers	4
<b>4</b>	<b>Information Theory</b>	<b>5</b>
<b>5</b>	<b>Bayesian</b>	<b>5</b>
<b>6</b>	<b>Optimal Stopping Theory</b>	<b>5</b>

---

## 1 Conventions

### Math Notation

## 2 Distributions

### 2.1 Gaussians

#### 2.1.1 Basics

1. To start with, *memorize* that

$$\int_{-\infty}^{\infty} dx e^{-x^2} = \pi^{1/2} \quad (2.1)$$

2. Next, anything multiplying the  $x^2$  in the integrand is present in inverse under the square root.

$$\int_{-\infty}^{\infty} dx e^{-\text{stuff } x^2} = \left( \frac{\pi}{\text{stuff}} \right)^{1/2} \quad (2.2)$$

so, for example:

$$\int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2} = \left( \frac{2\pi}{a} \right)^{1/2} \quad (2.3)$$

3. The traditional Gaussian pdf has  $a = 1/\sigma^2$ , and is easily seen to be

$$\mathcal{N}(x|0, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}x^2} \quad (2.4)$$

### 2.1.2 Differentiation moment trick

By differentiating Eq. 2.3 wrt  $a$ , we obtain an expression for integrals of the form  $\int_{-\infty}^{\infty} dx x^{2n} e^{-\frac{1}{2}ax^2}$ , with  $n \in \mathbb{Z}^+$ .

*e.g.* for  $n = 1$ :

$$-2 \frac{d}{da} \int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2} = \int_{-\infty}^{\infty} dx x^2 e^{-\frac{1}{2}ax^2} = -2 \frac{d}{da} \left( \frac{2\pi}{a} \right)^{1/2} = \left( \frac{2\pi}{a} \right)^{1/2} \frac{1}{a} \quad (2.5)$$

For  $n = 2$ :

$$\left( -2 \frac{d}{da} \right)^2 \int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2} = \int_{-\infty}^{\infty} dx x^4 e^{-\frac{1}{2}ax^2} = \left( -2 \frac{d}{da} \right)^2 \left( \frac{2\pi}{a} \right)^{1/2} = \left( \frac{2\pi}{a} \right)^{1/2} \frac{1}{a} \frac{3}{a} \quad (2.6)$$

Generally:

$$\int_{-\infty}^{\infty} dx x^{2n} e^{-\frac{1}{2}ax^2} = \left( \frac{2\pi}{a} \right)^{1/2} \frac{1}{a^n} (2n-1)(2n-3) \cdots 5 \cdot 3 \cdot 1 \quad (2.7)$$

We thus obtain an expression for the expectation value of  $x^{2n}$  under the Gaussian distribution:

$$\langle x^{2n} \rangle = \frac{\int_{-\infty}^{\infty} dx x^{2n} e^{-\frac{1}{2}ax^2}}{\int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2}} = \frac{1}{a^n} (2n-1)(2n-3) \cdots 5 \cdot 3 \cdot 1 \quad (2.8)$$

### 2.1.3 Gaussian with Linear Term

To evaluate integrals of the form

$$\int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2 + Jx}, \quad (2.9)$$

first complete the square in the exponent

$$-\frac{a}{2}x^2 + Jx = -\frac{a}{2} \left( x^2 - \frac{2Jx}{a} \right) = -\frac{a}{2} \left( x - \frac{J}{a} \right)^2 + \frac{J^2}{2a} \quad (2.10)$$

which gives

$$\int_{-\infty}^{\infty} dx e^{-\frac{1}{2}ax^2+Jx} = \int_{-\infty}^{\infty} dx e^{-\frac{1}{2}a(x-J/a)} e^{J^2/2a} = \left(\frac{2\pi}{a}\right)^{1/2} e^{J^2/2a} \quad (2.11)$$

where the integral is done by shifting  $x \rightarrow x + J/a$  (or noting that the infinite integral of a Gaussian is independent of its mean.)

By differentiating this expression wrt  $J$  repeatedly, and finally setting  $J = 0$ , we obtain another way of deriving the moments of the Gaussian, Eq. (2.8). This motivates the introduction of the **moment generating function**: given a pdf  $p(x)$ , the moment generating function is

$$\psi_x(J) = \mathbb{E}_x [e^{Jx}] = \int_{-\infty}^{\infty} dx e^{Jx} p(x) \quad (2.12)$$

which satisfies

$$\langle x^n \rangle = \left. \frac{d^n \psi_x(J)}{dJ^n} \right|_{J=0} \quad (2.13)$$

**TODO: Finish. Figure out clear way to include normalization factor of pdf in exposition**

## 2.1.4 Multivariate Gaussians

Promoting  $a$  to a real  $N \times N$  symmetric matrix  $\mathbf{A}$ , and  $x$  and  $J$  to a  $N$ -dim vectors  $\vec{x}$  and  $\vec{J}$  with components  $x_i$  and  $J_i$ , we have the multivariate Gaussian integral

$$\prod_{i=1}^N \left( \int_{-\infty}^{\infty} dx_i \right) e^{-\frac{1}{2}\vec{x}^T \mathbf{A} \vec{x} + \vec{J}^T \vec{x}} = \left( \frac{(2\pi)^N}{|\mathbf{A}|} \right)^{1/2} e^{\frac{1}{2}\vec{J}^T \mathbf{A}^{-1} \vec{J}} \quad (2.14)$$

**TODO: finish —**

A detailed walkthrough of multivariate Gaussian integrals is in [viXra:1404.0026](#).

## 2.2 Bernoulli

For  $x \in \{0, 1\}$ , Bernoulli dist parameterized by  $\mu$ , with

$$p(x; \mu) = \mu^x (1 - \mu)^{1-x} \quad (2.15)$$

## 3 Prob and stats

### 3.1 The Rules of Probability

- **Product Rule:**  $p(x, y) = p(x|y)p(y) = p(y|x)p(x)$
- **Sum Rule:**  $p(x) = \sum_y p(x, y) = \sum_y p(x|y)p(y)$

### 3.2 Bayes' Rule

Using  $p(y|x)p(x) = p(x, y) = p(x|y)p(y)$ , we have

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} = \frac{p(x|y)p(y)}{\sum_y p(x|y)p(y)} \quad (3.1)$$

### 3.3 Expectation and Variance

- **Expectations of sum of variables add:**

If  $X_1, \dots, X_n$  are random variables, and  $a_1, \dots, a_n$  are constants, then

$$\mathbb{E} \left( \sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i \mathbb{E}(X_i) \quad (3.2)$$

- **Variances of sum of independent variables add:**

If  $X_1, \dots, X_n$  are *independent* random variables, and  $a_1, \dots, a_n$  are constants, then

$$\text{Var} \left( \sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) \quad (3.3)$$

- **Variances of sum of (dependent) variables:**

If  $X$  and  $Y$  are random variables, then

$$\begin{aligned} \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ \text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y) \end{aligned} \quad (3.4)$$

### 3.4 Central Limit Theorem

Let  $S_N = \sum_{i=1}^N X_i$ , where each  $X_i$  is **iid** with mean  $\mu$  and variance  $\sigma^2$ . Then, as  $N \rightarrow \infty$ , the pdf of  $S_N$  approaches a normal distribution:

$$p(S_N = s) = \frac{1}{(2\pi N \sigma^2)^{1/2}} \exp \left[ -\frac{(s - N\mu)^2}{2N\sigma^2} \right] \quad (3.5)$$

NB the factors of  $N$  in the pdf, which make the pdf mean/variance equal to  $N$  times the original mean/variance (*i.e.* means and variances of independent variables add; see section 3.3.)

### 3.5 (Weak) Law of Large Numbers

Let  $X_1, \dots, X_n$  be **iid**, and  $\mu = \mathbb{E}(X_1)$ <sup>1</sup>. Defining the *sample mean* as  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ , the WLLN states that  $\bar{X}_n$  converges in probability to  $\mu$ .

---

<sup>1</sup> $\mu = \mathbb{E}(X_1) = \mathbb{E}(X_i)$  for any  $1 \leq i \leq n$

## 4 Information Theory

**KL divergence:**

$$\begin{aligned} KL[p(x)||q(x)] &= \sum_{x_i} p(x_i) \log \left( \frac{p(x_i)}{q(x_i)} \right) = - \sum_{x_i} p(x_i) \log \left( \frac{q(x_i)}{p(x_i)} \right) \\ &= - \sum_{x_i} p(x_i) \log q(x_i) + \sum_{x_i} p(x_i) \log p(x_i) \\ &= H(p, q) - H(p) \end{aligned} \tag{4.1}$$

where  $H(p, q)$  is the cross entropy, and  $H(p)$  is the entropy.

## 5 Bayesian

## 6 Optimal Stopping Theory