Gaussian random vectors

- The Gamma function
- The χ^2 distribution
- Confidence ellipsoids
- Marginal density functions
- Degenerate Gaussian random vectors
- Changes of variables for random vectors
- Linear transformations of Gaussians

Random variables

- we have a *real-valued random variable* x with *probability density function* (pdf) so that

$$\textbf{Prob}(x \in [a,b]) = \int_a^b p(x)\,dx$$

- the *mean* or *expected value* of x is $\quad \mathbf{E}(x) = \int_{-\infty}^{\infty} xp(x)\,dx$
- the *variance* of x is $\quad \textbf{var}(x) = \mathbf{E}\big((x-\mu)^2\big) = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)\,dx$
- mean is linear, so for $a, b \in \mathbb{R}$
  - $\mathbf{E}(ax+b) = a\,\mathbf{E}(x) + b$
  - $\textbf{var}(ax+b) = a^2\,\textbf{var}(x)$

$\mathbb{E}[X]$ $\mathbb{E}[Y]$

$f_X(x) = \cdot e^{-\frac{(x-\mu)^2}{2}}$

Gaussian random variables

$$\mathbb{E}(x) = \int x f(x)\,dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\, x\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \mu$$

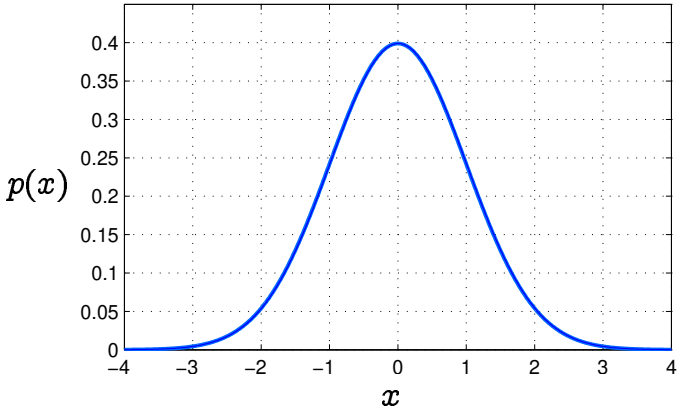
- x is *Gaussian* if it has *probability density function* (pdf) given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- write this as $x \sim \mathcal{N}(\mu, \sigma^2)$
- the *mean* or *expected value* of x is $\mathbf{E}(x) = \mu$
- the *variance* of x is $\mathbf{E}\big((x-\mu)^2\big) = \sigma^2$

Gaussian random variables

pdf for $x \sim \mathcal{N}(0, 1)$ is



- p is symmetric about the mean
- decays very fast; but $p(x) > 0$ for all x

Computing probabilities for Gaussian random variables

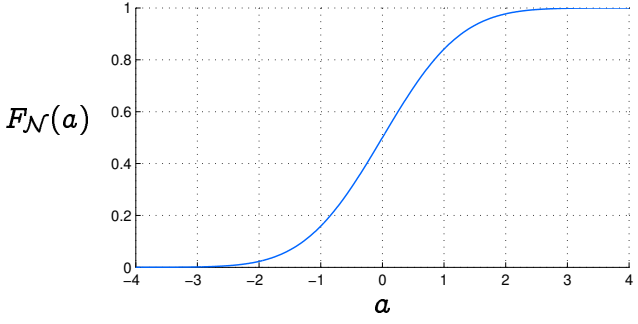
- the *error function* is

$$\mathbf{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

- the Gaussian *cumulative distribution function* (CDF) is

$$F_{\mathcal{N}}(a) = \mathbf{Prob}(x \leq a) = \frac{1}{2} + \frac{1}{2}\,\mathbf{erf}\left(\frac{a-\mu}{\sigma\sqrt{2}}\right)$$

- plot shows $\mu = 0$ and $\sigma = 1$,



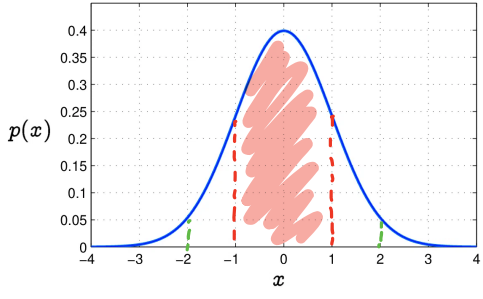
Computing probabilities for Gaussian random variables

- for $x \sim \mathcal{N}(0, \sigma^2)$ we have for $a \geq 0$

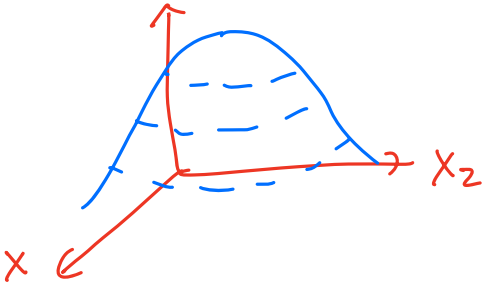
$$\textbf{Prob}(x \in [-a, a]) = \textbf{erf}\left(\frac{a}{\sigma\sqrt{2}}\right)$$

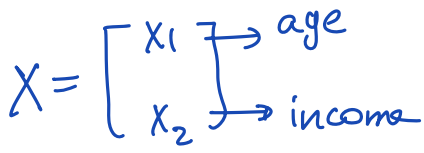
- some particular values:
  - $\textbf{Prob}(x \in [-\sigma, \sigma]) \approx 0.68$
  - $\textbf{Prob}(x \in [-2\sigma, 2\sigma]) \approx 0.9545$
  - $\textbf{Prob}(x \in [-3\sigma, 3\sigma]) \approx 0.9973$

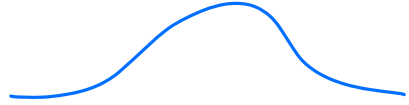
$\sigma = 1$

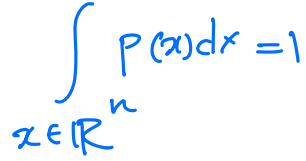


Continuous random vectors









- suppose $\mathbb{R}^n$-valued random vector x has probability density function $p^x : \mathbb{R}^n \to \mathbb{R}$.
- the *mean* or *expected value* of x is

$$\mathbf{E}(x) = \int_{\mathbb{R}^n} x p^x(x)\, dx$$

263 grades

X_1 \ X_2	A (3)	B (2)	C (1)	
undergrad ← 0	0.2	0.25	0.1	0.55
grad ← 1	0.25	0.15	0.05	0.45
	0.45	0.4	0.15	

$$P(\text{grade} \mid \text{undergrad}) = \left[\frac{0.2}{0.55}, \frac{0.25}{0.55}, \frac{0.1}{0.55}\right]$$

$$E[X] = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}$$

Covariance

$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \rightarrow Cov(X) = \begin{bmatrix} \cdots & Cov(X_i, X_j) & \cdots \end{bmatrix}$ (i, j) $\rightarrow \mathbb{E}((X_i - \mu_i)(X_j - \mu_j))$, $n \times n$

- the *covariance* of x is

$x^T A A^T x = (A^T x)^T (A^T x) = \|A^T x\|^2 \geq 0$

$$\mathbf{cov}(x) = \mathbf{E}\big((x-\mu)(x-\mu)^T\big) = \int_{\mathbb{R}^n} (x-\mu)(x-\mu)^T p^x(x)\, dx$$

($n \times 1$, $1 \times n$)

- We'll often denote the covariance by $\Sigma = \mathbf{cov}(x)$
- Σ is symmetric and positive semidefinite
- Σ_{ii} is the covariance of the i'th component x_i

$$\Sigma_{ii} = \mathbf{cov}(x_i) = \mathbb{E}\left((X_i - \mu_i)^2\right) = Var(X_i)$$

Affine transformations

- suppose x is an $\mathbb{R}^n$-valued random vector
- let y be an *affine function* of x, given by $y = Ax + b$
- the mean of y is the same affine function of the mean of x

$$\mathbf{E}(y) = A\,\mathbf{E}(x) + b$$

- the covariance of y is a linear function of the covariance of x

$$\mathbf{cov}(y) = A\,\mathbf{cov}(x)A^{\mathsf{T}}$$

$$\mathrm{Cov}(y) = \mathbb{E}\left((\underbrace{y - \mathbb{E}y}_{Ax})(\underbrace{y - \mathbb{E}y}_{(Ax)^{\mathsf{T}}})^{\mathsf{T}}\right) = \mathbb{E}\left(Axx^{\mathsf{T}}A^{\mathsf{T}}\right) = A\,\underbrace{\mathbb{E}(xx^{\mathsf{T}})}_{\mathrm{cov}(x)}A^{\mathsf{T}}$$

Affine transformations of random vectors

- for the mean, we have

$$\begin{aligned}\mathbf{E}\, y &= \mathbf{E}(Ax + b) \\ &= \int_{\mathbb{R}^n} (Ax + b)p(x)\,dx \\ &= b + A\int_{\mathbb{R}^n} xp(x)\,dx \\ &= b + A\,\mathbf{E}\, x\end{aligned}$$

- and for the covariance

$$\begin{aligned}\mathbf{cov}(y) &= \mathbf{E}\big((y - \mathbf{E}\, y)(y - \mathbf{E}\, y)^T\big) \\ &= \mathbf{E}\big(A(x - \mathbf{E}\, x)(x - \mathbf{E}\, x)^T A^T\big) \\ &= A\,\mathbf{E}\big((x - \mathbf{E}\, x)(x - \mathbf{E}\, x)^T\big)A^T \\ &= A\,\mathbf{cov}(x)A^T\end{aligned}$$

Mean-square deviation

- Suppose x is an $\mathbb{R}^n$-valued random variable, with mean μ.
- The *mean square deviation from the mean* is given by

$$\mathbf{E}\big(\|x-\mu\|^2\big) = \mathbf{trace}\,\mathbf{cov}(x)$$

- because

$$\begin{aligned}
\mathbf{E}\big(\|x-\mu\|^2\big) &= \mathbf{E}\big((x-\mu)^T(x-\mu)\big) \\
&= \mathbf{E}\,\mathbf{trace}\big((x-\mu)^T(x-\mu)\big) \\
&= \mathbf{E}\,\mathbf{trace}\big((x-\mu)(x-\mu)^T\big) && \text{since } \mathbf{trace}(AB) = \mathbf{trace}(BA) \\
&= \mathbf{trace}\,\mathbf{E}\big((x-\mu)(x-\mu)^T\big) && \text{since } \mathbf{E}\,Ax = A\,\mathbf{E}\,x
\end{aligned}$$

$$\mathbb{E}\left[\left\|\begin{bmatrix} x_1-\mu_1 \\ \vdots \\ x_n-\mu_n \end{bmatrix}\right\|^2\right] = \mathbb{E}\left((x_1-\mu_1)^2+\cdots+(x_n-\mu_n)^2\right)$$

$$= \mathbb{E}\left((x_1-\mu_1)^2\right)+\cdots+\mathbb{E}\left((x_n-\mu_n)^2\right)$$

$$\sigma_1^2 + \cdots + \sigma_n^2$$

The mean-variance decomposition

$$\mathrm{Var}(X) = \mathbb{E}\left((X-\mu)^2\right)$$
$$= \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right)$$
$$= \mathbb{E}(X^2) - 2\mu\,\underbrace{\mathbb{E}(X)}_{\mu} + \mu^2$$
$$= \mathbb{E}(X^2) - \mu^2$$

$$\mathbb{E}(X^2) = \mathrm{Var}(X) + (\mathbb{E}X)^2$$

- the *mean square* of a random variable x is

$$\mathbf{E}\left(||x||^2\right) = \mathbf{trace}\big(\mathbf{cov}(x)\big) + ||\mathbf{E}\,x||^2$$

- this holds because

$$\begin{aligned}\mathbf{E}\left(||x||^2\right) &= \mathbf{E}\left(||x-\mu+\mu||^2\right)\\ &= \mathbf{E}\left(||x-\mu||^2 + 2\mu^T(x-\mu) + ||\mu||^2\right)\\ &= \mathbf{E}\left(||x-\mu||^2\right) + 2\mu^T\,\mathbf{E}(x-\mu) + ||\mu||^2\end{aligned}$$

- called the *mean-variance decomposition*

Correlation coefficient

$$\frac{\Sigma_{ij}^2}{\Sigma_{ii}\Sigma_{jj}} \leq 1$$

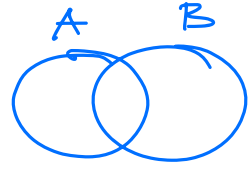
σ_i = standard deviation = $\sqrt{\text{variance}} = \sqrt{\sigma_i^2}$

- let $\Sigma = \mathbf{cov}(x)$. The *correlation coefficient* of x_i and x_j is

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = \frac{\mathrm{Cov}(X_i, X_j)}{\sigma_i \sigma_j}$$

- Since $\Sigma \geq 0$, we have $|\rho_{ij}| \leq 1$
- If $\rho_{ij} = 0$ then x_i and x_j are called *uncorrelated*.

indep ⇔ uncorrelated

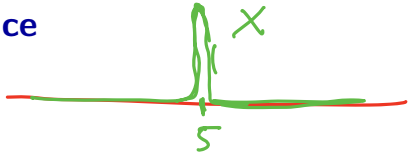


$P(A \cap B) = P(A)P(B)$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

indep → $P(A \cap B) = P(A)P(B)$ → $P(A|B) = P(A)$

Correlation and covariance



- the *correlation matrix* of random vector x is

$$\mathbf{corr}(x) = \mathbf{E}(xx^T)$$

- not to be confused with the correlation coefficient!
- If $\mathbf{E}\,x = 0$ then $\mathbf{corr}(x) = \mathbf{cov}(x)$
- The *mean square* of x is $\mathbf{E}\left(||x||^2\right) = \mathbf{trace}\,\mathbf{corr}(x)$
- The *correlation-covariance decomposition* is

$$\mathbf{corr}(x) = \mathbf{cov}(x) + (\mathbf{E}\,x)(\mathbf{E}\,x^T)$$

- same approach as the mean-variance formula

Gaussian random vectors

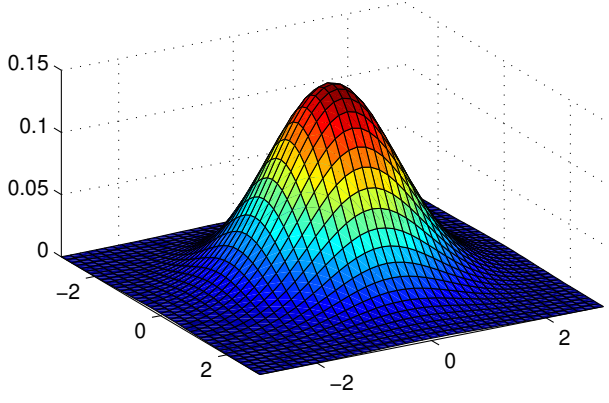
$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

- the $\mathbb{R}^n$-valued random variable x is called *Gaussian* if it has pdf

$$p^x(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

RV, dummy, $x^T A x$

- write this as $x \sim \mathcal{N}(\mu, \Sigma)$, here $\Sigma = \Sigma^T$ and $\Sigma > 0$



Gaussian random vectors

- suppose $x \sim \mathcal{N}(\mu, \Sigma)$. Then the mean of x is

$$\mathbf{E}\, x = \mu$$

- and the covariance of x is

$$\mathbf{cov}(x) = \Sigma$$

Ellipsoids

lued random variable x is called *Gaussian* if it has pdf

$$p^{(x)}(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

RV, dummy

as $x \sim \mathcal{N}(\mu, \Sigma)$, here $\Sigma = \Sigma^T$ and $\Sigma > 0$

- the Gaussian pdf is constant on the surface of the ellipsoids

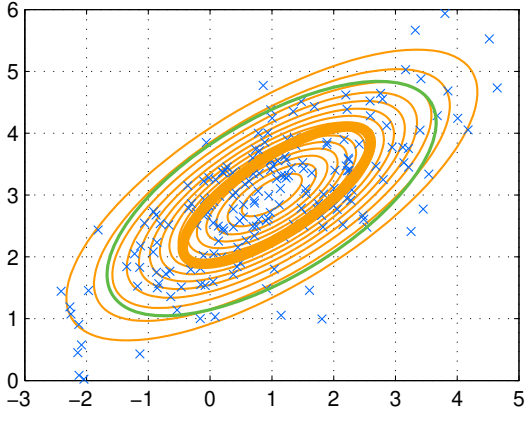
$$S_\alpha = \left\{ x \in \mathbb{R}^n \;\middle|\; (x-\mu)^T \Sigma^{-1} (x-\mu) \leq \alpha \right\}$$

- center is at μ, semiaxis lengths are $\sqrt{\alpha \lambda_i(\Sigma)}$.
- example has

$$\mu = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

contours at $p(x) = 0.01, 0.02, \ldots$



Gamma function

- the *gamma function* is

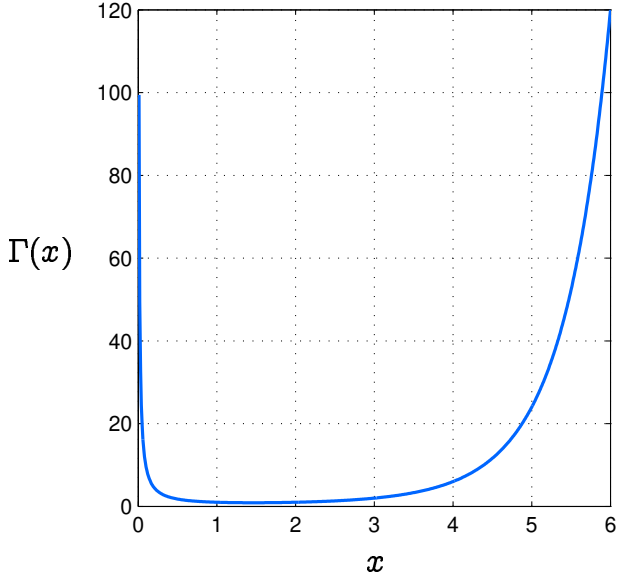
$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \qquad \text{for } x > 0$$

- for $x > 0$

$$\Gamma(x+1) = x\Gamma(x)$$

- $\Gamma(1) = 1$, so for integer $x > 1$

$$\Gamma(x) = (x-1)!$$

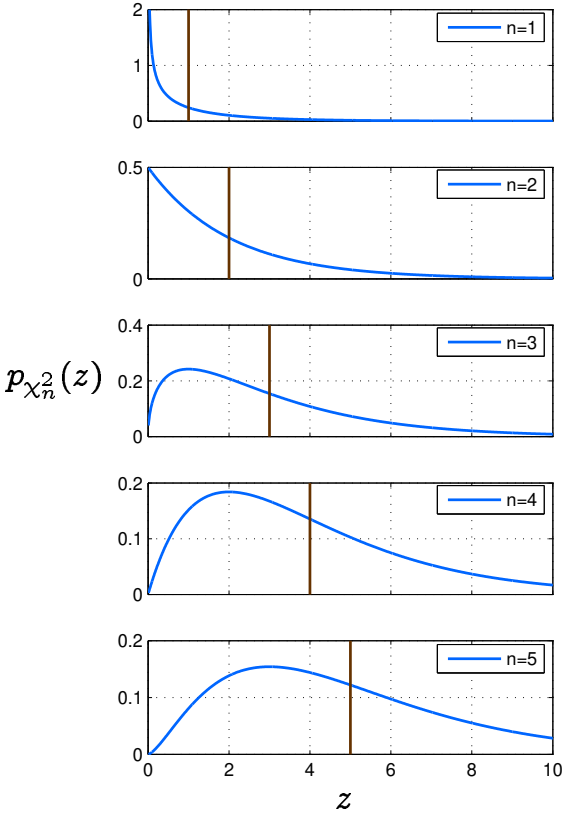


The χ^2 distribution

- the χ^2_n probability density function is

$$p_{\chi^2_n}(z) = \frac{1}{2^{\frac{n}{2}}\Gamma(n/2)} z^{\frac{n}{2}-1} e^{-\frac{z}{2}}$$

- A family of pdfs, one for each $n > 0$
- If $z \sim \chi^2_n$, then $\mathbf{E}\, z = n$



Gaussian random vectors and confidence ellipsoids

- suppose x is Gaussian, i.e., $x \sim \mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$. Define the random variable

$$z = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

  which is a measure of the distance of x from μ

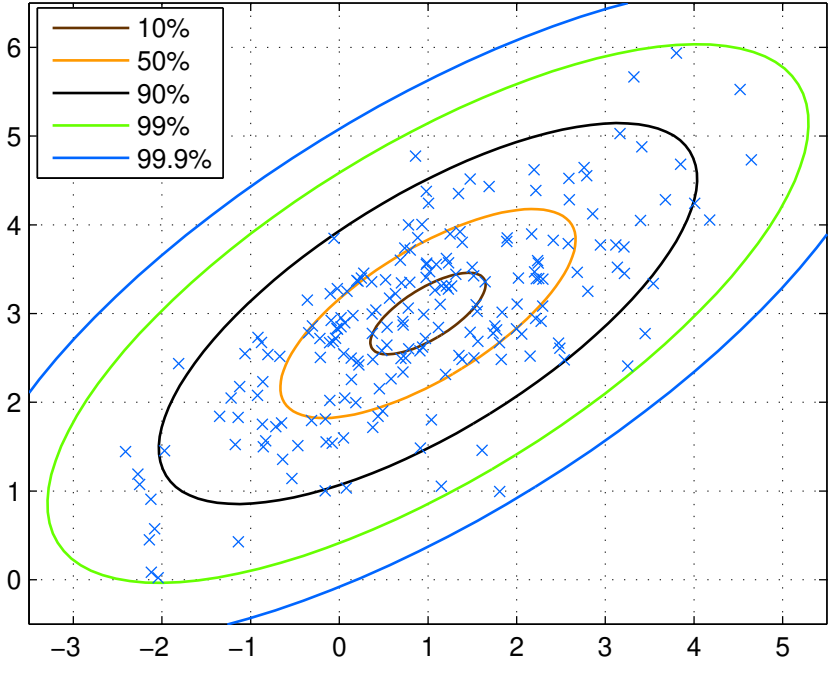
- z has a χ^2_n distribution
- Hence prob. that x lies in the ellipsoid $S_\alpha = \left\{ x \in \mathbb{R}^n \;\middle|\; (x - \mu)^T \Sigma^{-1} (x - \mu) \leq \alpha \right\}$

$$\textbf{Prob}(x \in S_\alpha) = F_{\chi^2_n}(\alpha)$$

- for example $F_{\chi^2_n}(\alpha) \approx \begin{cases} \frac{1}{2} & \text{if } \alpha = n \\ 0.9 & \text{if } \alpha = n + 2\sqrt{n} \end{cases}$ 90% confidence ellipsoid

Confidence ellipsoids

The plot shows the confidence ellipsoids and 200 sample points.



Marginal probability density functions

- suppose x is an RV with pdf $p^x : \mathbb{R}^n \to \mathbb{R}$, and $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, where $x_1 \in \mathbb{R}^r$.
- define the *marginal pdf* of x_1 to be the function p^{x_1} such that

$$\textbf{Prob}(x_1 \in W) = \int_W p^{x_1}(z)\, dz \qquad \text{for all } W \subset R^r$$

- we also know that

$$\textbf{Prob}(x_1 \in W) = \int_W \int_{x_2 \in \mathbb{R}^{n-r}} p^x(x_1, x_2)\, dx_2\, dx_1$$

- since these are equal, we have

$$p^{x_1}(x_1) = \int_{x_2 \in \mathbb{R}^{n-r}} p^x(x_1, x_2)\, dx_2$$

$P(x_1) = \sum_{x_2} P(x_1, x_2)$

The marginal pdf of a Gaussian

- suppose $x \sim \mathcal{N}(\mu, \Sigma)$, and

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- let's look at the component x_1
- Since $x_1 = \begin{bmatrix} I & 0 \end{bmatrix} x$, we have the mean

$$\mathbf{E}\, x_1 = \begin{bmatrix} I & 0 \end{bmatrix} \mu = \mu_1$$

  and also the covariance

$$\mathbf{cov}(x_1) = \begin{bmatrix} I & 0 \end{bmatrix} \Sigma \begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_{11}$$

- In fact, the random variable x_1 is *Gaussian*; this is not obvious

$X_1, X_2 \sim N(0,1)$

$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N(\mu, \Sigma)$

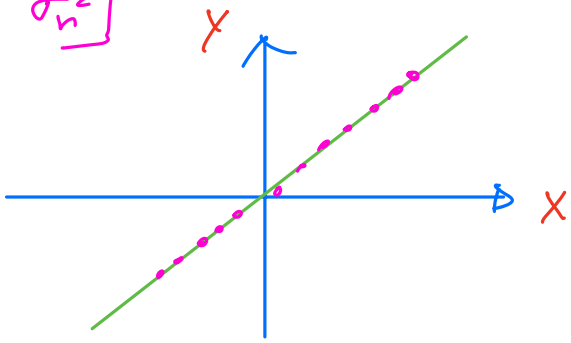
X_1, X_2 iid $\rightarrow \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ 0 & \cdots & \sigma_n^2 \end{bmatrix}$

$X \sim N(0,1)$

$S \in \{+1, -1\}$

$Y = SX$

$Y \sim N(0,1)$



Proof: the marginal pdf of a Gaussian

- assume for convenience that $\mathbf{E}\, x = 0$. The marginal pdf of x_1 is

$$p^{x_1}(x_1) = \int_{x_2} c_1 \exp\left(-\frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) dx_2$$

- we have, by the completion of squares formula

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} I & -\Sigma_{11}^{-1}\Sigma_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix}$$

- and so, setting $S = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^T \Sigma_{11}^{-1} x_1 \ + \ (x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)^T S^{-1} (x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)$$

Proof: the marginal pdf of a Gaussian

- hence we have

$$p^{x_1}(x_1) = c_1 \exp\left(-\frac{1}{2} x_1^T \Sigma_{11}^{-1} x_1\right) \int_{x_2} \exp\left(-\frac{1}{2}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)^T S^{-1}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1)\right) dx_2$$

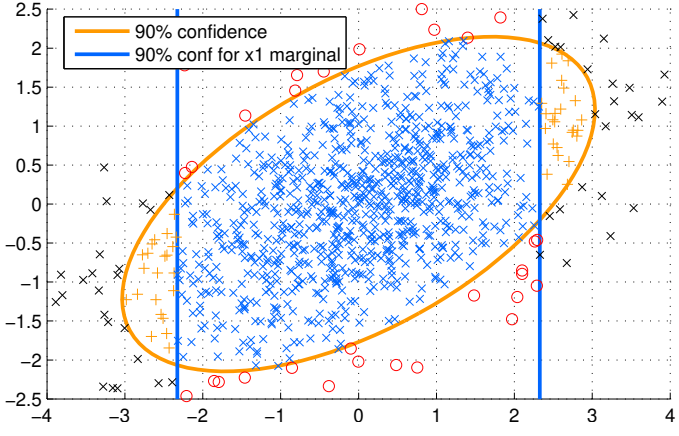
$$= c_2 \exp\left(-\frac{1}{2} x_1^T \Sigma_{11}^{-1} x_1\right)$$

- now c_2 is determined, because $\int p^{x_1}(z)\, dz = 1$, so we don't need to calculate it explicitly.
- therefore, if $x \sim \mathcal{N}(0, \Sigma)$ the marginal pdf of x_1 is *Gaussian*, and

$$x_1 \sim \mathcal{N}(0, \Sigma_{11})$$

Example: marginal pdf for Gaussians

- Suppose $\Sigma = \begin{bmatrix} 2 & 0.8 \\ 0.8 & 1 \end{bmatrix}$ and $x \sim \mathcal{N}(0, \Sigma)$. A simulation of 1000 points is below
- all blue and orange points (908) are within 90% confidence ellipsoid for x
- all blue and red points (899) are within 90% confidence interval for x_1



Degenerate Gaussian random vectors

$p(x) \sim \exp\left(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)\right)$

Real / sym. $A = Q\Lambda Q^T$

- it's convenient to allow Σ singular, but still $\Sigma = \Sigma^T$ and $\Sigma \geq 0$
  this means that in some directions, *x is not random at all*
- obviously density formula does not hold; instead write

$$\Sigma = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T$$

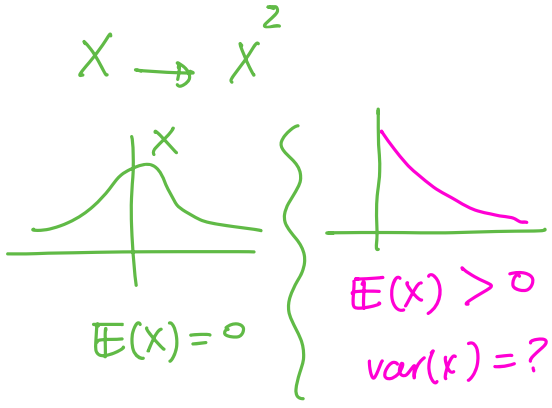
  where $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ is orthogonal, and $\Sigma_1 > 0$

  columns of Q_1 are orthonormal basis for **range**(Σ)
  columns of Q_2 are orthonormal basis for **null**(Σ)

$Q^T x = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}^T x = \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} x = \begin{bmatrix} Q_1^T x \\ Q_2^T x \end{bmatrix} \to \begin{matrix} z \\ w \end{matrix}$

- let $\begin{bmatrix} z \\ w \end{bmatrix} = Q^T x$; then
  - $z \sim \mathcal{N}(Q_1^T \mu, \Sigma_1)$ is non-degenerate Gaussian
  - $w = Q_2^T \mu$ is not random

Changes of variables for random vectors



- suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, and $h : \mathbb{R}^n \to \mathbb{R}^n$ satisfies
  - h is one-to-one and onto; i.e., h is invertible
  - Both h and h^{-1} are differentiable, with continuous derivative
- the derivative of h at x is $Dh(x)$, the *Jacobian* matrix

$$\big(Dh(x)\big)_{ij} = \frac{\partial h_i}{\partial x_j}(x)$$

- then for any $A \subset \mathbb{R}^n$

$$\int_{h(A)} f(x)\,dx = \int_A f\big(h(y)\big)\,\big|\mathbf{det}\; Dh(y)\big|\,dy$$

Changes of variables for random vectors

- suppose x is an $\mathbb{R}^n$-valued random vector, and $y = g(x)$, where g is invertible, and g and g^{-1} are continuously differentiable. Then

$$p^y(y) = \frac{p^x\big(g^{-1}(y)\big)}{\left|\mathbf{det}(Dg)\big(g^{-1}(y)\big)\right|}$$

- this holds because

$$\begin{aligned}
\mathbf{Prob}(y \in A) &= \int_A p^y(y)\,dy \\
&= \int_{g^{-1}(A)} p^x(x)\,dx \\
&= \int_A \frac{p^x\big(g^{-1}(y)\big)}{\left|\mathbf{det}(Dg)\big(g^{-1}(y)\big)\right|}\,dy
\end{aligned}$$

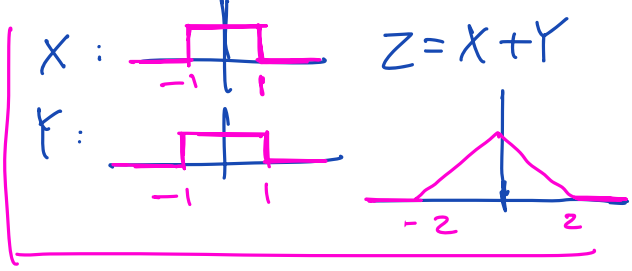
where $D(g^{-1})(y) = \Big((Dg)\big(g^{-1}(y)\big)\Big)^{-1}$

Example: linear transformations

- consider $y = Ax + b$, where $A \in \mathbb{R}^{n \times n}$ is *invertible*. Then

$$p^y(y) = \frac{p^x\left(A^{-1}(y-b)\right)}{|\det A|}$$

Linear transformations of Gaussians



- fundamental result: a linear function of a Gaussian random vector is a Gaussian random vector
- suppose $x \sim \mathcal{N}(\mu_x, \Sigma_x)$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Consider the linear function of x

$$y = Ax + b$$

- we already know how means and covariances transform; we have

$$\mathbf{E}(y) = A\,\mathbf{E}\,x + b \qquad\qquad \mathbf{cov}(y) = A\,\mathbf{cov}(x)A^T$$

- additional fact is that y is *Gaussian*

Linear transformations of Gaussians

- to show this, first suppose $A \in \mathbb{R}^{n \times n}$ is *invertible*. Let $\mu_y = A\mu_x + b$ and $\Sigma_y = A\Sigma_x A^T$.
- we know

$$p^x(x) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma_x)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma_x^{-1}(x-\mu)\right)$$

- so

$$\begin{aligned} p^y(y) &= \frac{p^x\left(A^{-1}(y-b)\right)}{|\det A|} \\ &= \frac{1}{|\det A|(2\pi)^{\frac{n}{2}} (\det \Sigma_x)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-b-A\mu_x)^T (A^{-1})^T \Sigma_x^{-1} A^{-1}(y-b-A\mu_x)\right) \\ &= \frac{1}{(2\pi)^{\frac{n}{2}} (\det \Sigma_y)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-\mu_y)^T \Sigma_y^{-1}(y-\mu_y)\right) \end{aligned}$$

Non-invertible linear transformations of Gaussians

- suppose $A \in \mathbb{R}^{m \times n}$, and $y = Ax$ where $x \sim \mathcal{N}(0, \Sigma_x)$. The SVD of A is

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

- so we decompose the map into

$$y = Uw \qquad \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \Sigma z \qquad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = V^T x$$

Non-invertible linear transformations of Gaussians

- since V is invertible, we know $z \sim \mathcal{N}(0, \Sigma_z)$, where

$$\Sigma_z = \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \Sigma_x \begin{bmatrix} V_1 & V_2 \end{bmatrix}$$

- we know z is Gaussian, hence the marginal z_1 is Gaussian

$$z_1 \sim \mathcal{N}(0, V_1^T \Sigma_x V_1)$$

- also $w_2 = 0$, and since Σ_1 is invertible, w_1 is Gaussian

$$w_1 \sim \mathcal{N}(0, \Sigma_1 V_1^T \Sigma_x V_1 \Sigma_1)$$

- since $w = U^T y$, we have y is a *degenerate Gaussian random vector* where
  - $w_1 = U_1^T y$ are the components of y that are Gaussian
  - $w_2 = 0$ are the components of y that are not random

Full-rank case

- when **range**$(A) = \mathbb{R}^m$, i.e., A is full row rank, we have

$$y \sim \mathcal{N}(0, A\Sigma_x A^T)$$

- because the SVD of A is

$$A = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

- then $y = Uw_1$, and since U is invertible, we have

$$y \sim \mathcal{N}(0, U\Sigma_1 V_1^T \Sigma_x V_1 \Sigma_1 U^T)$$

Example: simulating Gaussian random vectors

- in many languages its easy to generate $x \sim \mathcal{N}(0, I)$
- to generate $y \sim \mathcal{N}(\mu, \Sigma)$, we can use

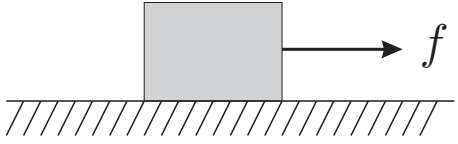
$$y = \Sigma^{\frac{1}{2}} x + \mu$$

- extremely useful for simulation

Example: Gaussian random force on mass

- x is the sequence of applied forces, so $f(t) = x_j$ for t in the interval $[j-1, j]$.
- y_1, y_2 are final position and velocity
- $y = Ax$ where $A = \begin{bmatrix} 9.5 & 8.5 & 7.5 & 6.5 & 5.5 & 4.5 & 3.5 & 2.5 & 1.5 & 0.5 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$
- suppose the forces are Gaussian, and the vector $x \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 2 & 1 & & & & & & & & \\ 1 & 2 & 1 & & & & & & & \\ & 1 & 2 & 1 & & & & & & \\ & & 1 & 2 & 1 & & & & & \\ & & & 1 & 2 & 1 & & & & \\ & & & & 1 & 2 & 1 & & & \\ & & & & & 1 & 2 & 1 & & \\ & & & & & & 1 & 2 & 1 & \\ & & & & & & & 1 & 2 & 1 \\ & & & & & & & & 1 & 2 \end{bmatrix}$$



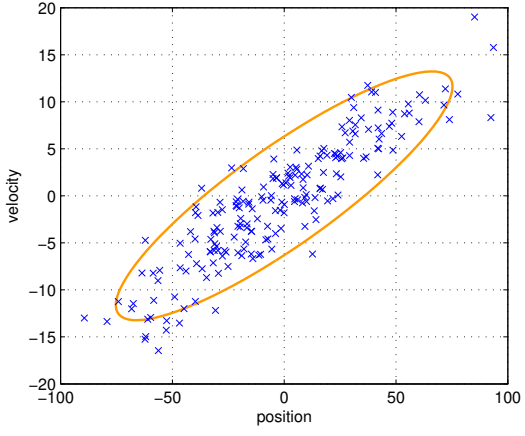
Example: Gaussian random force on mass

- the covariance of y is

$$\Sigma_y = A\Sigma A^T$$

- the 90% confidence ellipsoid is

$$\left\{ y \in \mathbb{R}^2 \mid y^T \Sigma_y^{-1} y \leq F_{\chi_n^2}^{-1}(0.9) \right\}$$



Components of a Gaussian random vector

- suppose $x \sim \mathcal{N}(0, \Sigma)$, and let $c \in \mathbb{R}^n$ be a unit vector
- let $y = c^T x$
- y is the component of x in the direction c
- y is Gaussian, with $\mathbf{E}\, y = 0$ and $\mathbf{cov}(y) = c^T \Sigma c$
- So $\mathbf{E}(y^2) = c^T \Sigma c$
- The unit vector c that minimizes $c^T \Sigma c$ is the eigenvector of Σ with the smallest eigenvalue. Then

$$\mathbf{E}(y^2) = \lambda_{\mathbf{min}}$$