ME 7247: Advanced Control Systems **Fall 2022–23**

Lecture 08: Random Variables and Random Vectors

Tuesday October 4, 2022

Lecturer: Laurent Lessard *Scribe: Cory Lafleur*

This lecture begins by describing random variables ($x \in \mathbb{R}$). The properties of the probability density function, the mean, and the variance are discussed along with some common distributions, normal and χ^2. Then, random vectors ($x \in \mathbb{R}^n$) are introduced along with the related probability density function, expected value, covariance, and multivariate distributions.

1 Random Variables

A random variable, $x \in \mathbb{R}$, does not have a definitive value. Instead, it may take on a variety of values associated with its probability density.

1.1 Probability Density Function

A function associated with a random variable such that $f : \mathbb{R} \to \mathbb{R}^+$. The higher the density under a particular region of the probability density function (pdf), the more likely that value is to occur.

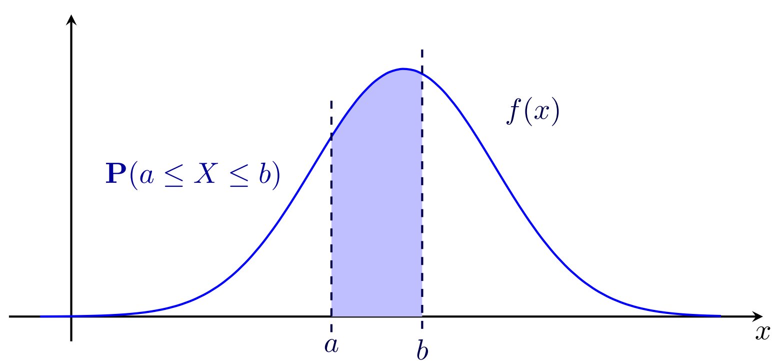

Figure 1: Example Probability Density Function

The probability is given by the area under the curve between two points. Therefore,

$$\mathbf{P}(a \le X \le b) = \int_a^b f(x)\,dx.$$

The properties of the probability density function are:

- Positivity: $f(x) \ge 0$ for all x.
- Integrates to one: $\int_{-\infty}^{\infty} f(x)\,dx = \int_{\mathbb{R}} f(x)\,dx = 1$

The cumulative density function (CDF) is defined as

$$F(x) := \int_{-\infty}^{x} f(t)\,dt.$$

Then we have $\lim_{x\to\infty} F(x) = 1$.

1.2 Numerical summaries

Mean (Expected Value) (μ):

$$\mathbf{E}(x) = \int_{\mathbb{R}} x f(x)\,dx. \tag{1}$$

Variance (σ^2):

$$\mathbf{Var}(x) = \mathbf{E}\left[(x - \mathbf{E}(x))^2\right]. \tag{2}$$

Standard Deviation (σ): the positive square root of variance.

1.2.1 Transformation Properties

Expected Value of a Linear Transformation

$$\mathbf{E}(ax+b) = a\mathbf{E}(x) + b.$$

where $a, b \in \mathbb{R}$ and x is a r.v.

Proof.

$$\begin{aligned}
\mathbf{E}(ax+b) &\overset{(1)}{=} \int_{\mathbb{R}} (ax+b) f(x)\,dx \\
&= a\int_{\mathbb{R}} x f(x)\,dx + b\int_{\mathbb{R}} f(x)\,dx \\
&= a\mathbf{E}(x) + b.
\end{aligned}$$

∎

Variance in Terms of Expected Value

$$\mathbf{Var}(x) = \mathbf{E}(x^2) - \mathbf{E}(x)^2. \tag{3}$$

Proof.

$$\begin{aligned}
\mathbf{Var}(x) &= \mathbf{E}[(x - \mathbf{E}(x))^2] && \text{from (2)} \\
&= \mathbf{E}(x^2 - 2x\mathbf{E}(x) + \mathbf{E}(x)^2) \\
&= \mathbf{E}(x^2) - 2\mathbf{E}(x\mathbf{E}(x)) + \mathbf{E}(x)^2 \\
&= \mathbf{E}(x^2) - 2\mathbf{E}(x)^2 + \mathbf{E}(x)^2 \\
&= \mathbf{E}(x^2) - \mathbf{E}(x)^2.
\end{aligned}$$

∎

Variance of a Linear Transformation

$$\mathbf{Var}(ax+b) = a^2\mathbf{Var}(x).$$

Proof.

$$\begin{aligned}\mathbf{Var}(ax+b) &= \mathbf{E}\left[((ax+b) - \mathbf{E}(ax+b))^2\right] \\ &= \mathbf{E}\left[(ax+b-a\mathbf{E}(x)-b)^2\right] \\ &= \mathbf{E}\left[a^2(x-\mathbf{E}(x))^2\right] \\ &= a^2\mathbf{E}\left[(x-\mathbf{E}(x))^2\right] \\ &= a^2\mathbf{Var}(x).\end{aligned}$$

■

1.3 Normal Distributions

$$x \sim \mathcal{N}(\mu,\, \sigma^2)$$

x is normally distributed with mean μ and variance σ^2.

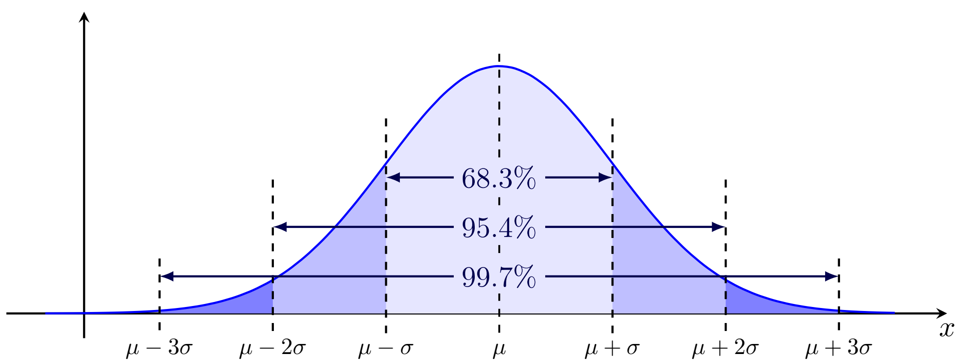

Figure 2: Normal Distribution

The equation of the probability density function of a normal distribution is

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

The expected value and variance are given by

$$\mathbf{E}(x) = \mu \quad \text{and} \quad \mathbf{Var}(x) = \sigma^2.$$

If $x \sim \mathcal{N}(\mu,\, \sigma^2)$ then $z = \frac{x-\mu}{\sigma} \sim \mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ is the standard normal. Some useful MATLAB commands:

- `normpdf(x,mu,sigma)` $= f(x)$, where $x \sim \mathcal{N}(\mu, \sigma^2)$
- `normcdf(x,mu,sigma)` $= F(x)$, where $x \sim \mathcal{N}(\mu, \sigma^2)$
- `normpdf(x)` $= f(x)$, where $x \sim \mathcal{N}(0,1)$

1.4 Chi Squared χ^2

If $w = z_1^2 + z_2^2 + \cdots + z_k^2$, where each $z_i \sim \mathcal{N}(0,1)$ is a standard normal, then $w \sim \chi_k^2$, where k is the *degrees of freedom*.

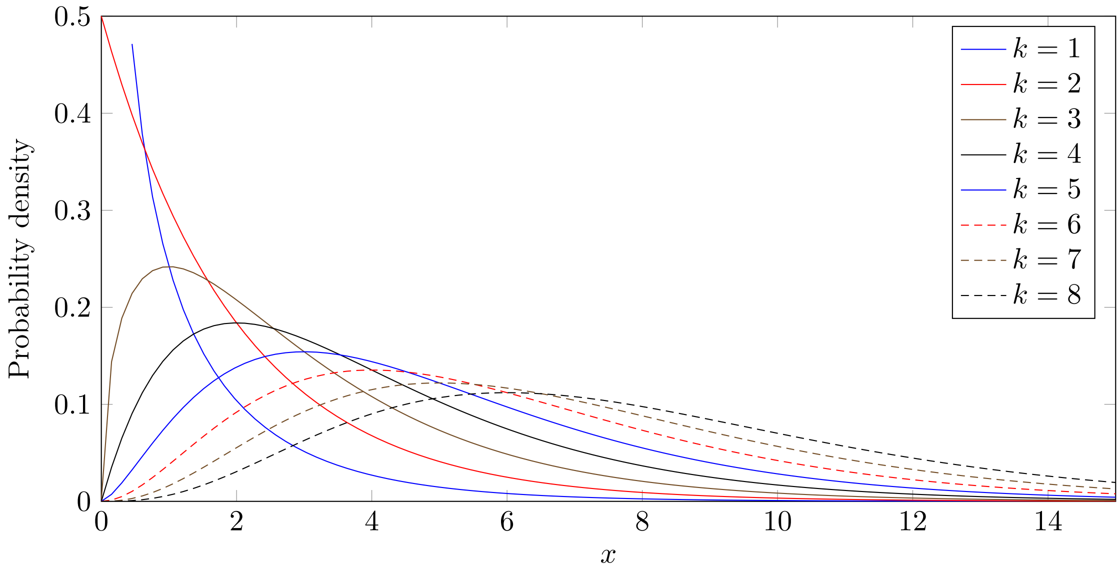

Figure 3: Plot of the PDF of χ_k^2 for different values of k.

The expected value and variance for $w \sim \chi_k^2$ are given by

$$\mathbf{E}(w) = k \quad \text{and} \quad \mathbf{Var}(w) = 2k.$$

As $k \to \infty$, we also have $\chi_k^2 \to \mathcal{N}(k, 2k)$.

2 Random Vectors

A random vector, $x \in \mathbb{R}^n$. We will again define the probability density function, summaries, and distributions for random vectors.

2.1 Probability Density Function

A function associated with a random vector such that $f : \mathbb{R}^n \to \mathbb{R}^+$.

The probability that x lies within the set S is given by

$$\mathbf{P}(x \in S) = \int_S f(x)\, dx.$$

The properties of the probability density function are:

- Positivity: $f(x) \geq 0$ for all x.
- Integrates to one: $\int \cdots \int_{\mathbb{R}^n} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int_{\mathbb{R}^n} f(x)\, dx = 1$

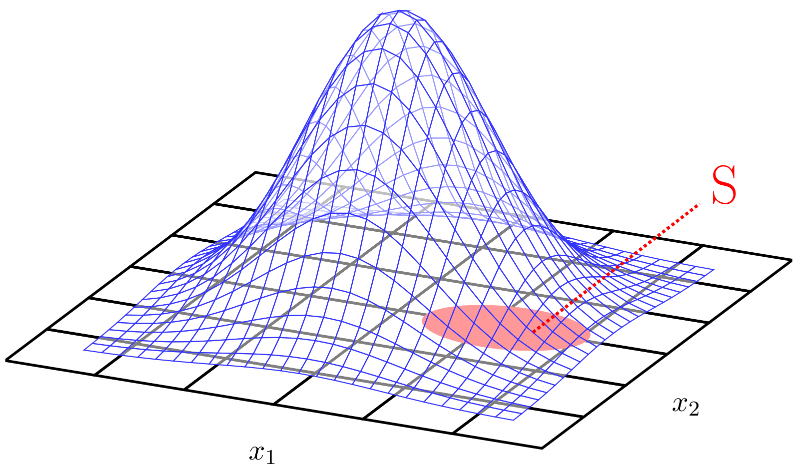

Figure 4: Probability Density Function in $\mathbb{R}^2$.

2.2 Summaries

Expected Value ($\mu \in R^n$):

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

Covariance ($\Sigma \in R^{n \times n}$):

$$\mathbf{Cov}(x) = \mathbf{E}\left[(x - \mathbf{E}\, x)(x - \mathbf{E}\, x)^{\mathsf{T}}\right]. \tag{4}$$

In $\mathbb{R}^2$

$$\mathbf{Cov}(x) = \mathbf{E}\begin{bmatrix} (x_1 - \mathbf{E}\, x_1)^2 & (x_1 - \mathbf{E}\, x_1)(x_2 - \mathbf{E}\, x_2) \\ (x_1 - \mathbf{E}\, x_1)(x_2 - \mathbf{E}\, x_2) & (x_2 - \mathbf{E}\, x_2)^2 \end{bmatrix}.$$

2.2.1 Properties

Linear Transformation

$$\mathbf{E}(Ax + b) = A\mathbf{E}(x) + b.$$

$$\mathbf{Cov}(Ax + b) = A\mathbf{Cov}(x)A^{\mathsf{T}}.$$

Proof.

$$\begin{aligned}
\mathbf{Cov}(Ax+b) &\overset{(4)}{=} \mathbf{E}\left[((Ax+b)-\mathbf{E}(Ax+b))((Ax+b)-\mathbf{E}(Ax+b))^{\mathsf{T}}\right] \\
&= \mathbf{E}\left[(Ax+b-A\mathbf{E}(x)-b)(Ax+b-A\mathbf{E}(x)-b)^{\mathsf{T}}\right] \\
&= \mathbf{E}\left[A(x-\mathbf{E}(x))(x-\mathbf{E}(x))^{\mathsf{T}}A^{\mathsf{T}}\right] \\
&= A\mathbf{E}\left[(x-\mathbf{E}(x))(x-\mathbf{E}(x))^{\mathsf{T}}\right]A^{\mathsf{T}} \\
&= A\mathbf{Cov}(x)A^{\mathsf{T}}.
\end{aligned}$$

■

Variance in Terms of Expected Value

$$\mathbf{Cov}(x) = \mathbf{E}\left(xx^{\mathsf{T}}\right) - \mathbf{E}(x)\mathbf{E}(x)^{\mathsf{T}}. \tag{5}$$

The proof of this fact is analogous to the proof of (3).

Positive definiteness. The covariance matrix is positive definite. That is, $\mathbf{Cov}(x) \succ 0$.

To review, there are two equivalent definitions of positive definiteness. A symmetric matrix $Q = Q^{\mathsf{T}}$ is positive definite ($Q \succ 0$) if either of the following equivalent properties hold.

(i) All eigenvalues of Q are positive: $\lambda_i > 0$.

(ii) All quadratic forms of Q are positive: $x^{\mathsf{T}}Qx > 0$ for all $x \neq 0$.

Also, we can define a "negative definite" matrix $Q \prec 0 \iff -Q \succ 0$. We also write that one matrix is larger than another in the *positive definite sense* by writing $Q \succ R \iff Q - R \succ 0$. If a matrix has at least one positive eigenvalue and at least one negative eigenvalue, we say that it is "indefinite". To prove positive definiteness, we use the quadratic form definition.

Proof.

$$\begin{aligned}
v^{\mathsf{T}}\mathbf{Cov}(x)v &= v^{\mathsf{T}}\mathbf{E}\left[(x-\mathbf{E}(x))(x-\mathbf{E}(x))^{\mathsf{T}}\right]v \\
&= \mathbf{E}\left[v^{\mathsf{T}}(x-\mathbf{E}(x))(x-\mathbf{E}(x))^{\mathsf{T}}v\right] \\
&= \mathbf{E}\left[\left(v^{\mathsf{T}}(x-\mathbf{E}(x))\right)^2\right]
\end{aligned}$$

This is the expected value of a quantity that is always nonnegative. The only way it can be zero for all $v \neq 0$ is if $x = \mathbf{E}(x)$. In other words, the random variable x needs to not be random at all. In this case, we actually have $\mathbf{Cov}(x) = 0$. If x is random, then the quadratic form is positive whenever $v \neq 0$, and therefore $\mathbf{Cov}(x) \succ 0$. ■

Expected Value of Quadratic Form If $Q = Q^\top$ and x is a random vector with $\mathbf{E}(x) = \mu$ and $\mathbf{Cov}(x) = \Sigma$, then we can evaluate the expected value of a quadratic form using the formula

$$\mathbf{E}\left(x^\top Q x\right) = \mu^\top Q \mu + \mathbf{tr}(Q\Sigma).$$

Proof. To prove this fact, we will use the notion of *trace* of a square matrix, which is just the sum of the diagonal entries of the matrix. If A and B are matrices such that AB and BA are both square, then the trace has the property $\mathbf{tr}(AB) = \mathbf{tr}(BA)$. Now compute:

$$\begin{aligned}
\mathbf{E}\left(x^\top Q x\right) &= \mathbf{E}\left(\mathbf{tr}(x^\top Q x)\right) && \text{since } a = \mathbf{tr}(a) \text{ for scalar } a. \\
&= \mathbf{E}\left(\mathbf{tr}(Qxx^\top)\right) && \text{using } \mathbf{tr}(AB) = \mathbf{tr}(BA) \\
&= \mathbf{tr}\left(\mathbf{E}(Qxx^\top))\right) && \text{trace and expectation commute} \\
&= \mathbf{tr}\left(Q\mathbf{E}(xx^\top))\right) && \text{linearity of expectation} \\
&= \mathbf{tr}\left(Q(\mathbf{Cov}(x) + \mathbf{E}(x)\mathbf{E}(x)^\top)\right) && \text{using (5)} \\
&= \mathbf{tr}\left(Q(\Sigma + \mu\mu^\top)\right) \\
&= \mathbf{tr}(Q\Sigma) + \mathbf{tr}\left(Q\mu\mu^\top\right) \\
&= \mu^\top Q \mu + \mathbf{tr}(Q\Sigma) && \text{using } \mathbf{tr}(AB) = \mathbf{tr}(BA) \text{ again}
\end{aligned}$$

■

As a sanity check, we can verify that the formula has the correct limiting behavior. First, if x is not random at all, then $\Sigma = 0$ and $x = \mu$, and we obtain $\mathbf{E}(x^\top Q x) = \mu^\top Q \mu$, as expected. If x has zero mean, then $\mu = 0$ and $\mathbf{E}(x^\top Q x) = \mathbf{tr}(Q\Sigma)$. So the quadratic form has a larger value when x has more variance.

2.3 Multivariate Normal Distributions

We write this as: $x \sim \mathcal{N}(\mu, \Sigma)$, where μ is the mean and Σ is the covariance.

The probability density function (pdf) of a Multivariate Gaussian Distribution is given by:

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

If $x \sim \mathcal{N}(\mu, \Sigma)$, then $z := \Sigma^{-\frac{1}{2}}(x - \mu)$ has a *standard normal distribution*, $z \sim \mathcal{N}(0, I)$. To find the matrix square root, take an eigenvalue decomposition: $\Sigma = U\Lambda U^\top$. Because $\Sigma \succ 0$ (it's a covariance matrix), the eigenvalue decomposition is the same as the singular value decomposition. The matrix square root is:

$$\Sigma^{\frac{1}{2}} = U \begin{bmatrix} \lambda_1^{\frac{1}{2}} & & \\ & \ddots & \\ & & \lambda_n^{\frac{1}{2}} \end{bmatrix} U^\top \tag{6}$$

Although a matrix has 2^n possible square roots (we can write $\pm\lambda_i^{1/2}$ for each i), there is only one that positive definite. The MATLAB command for computing this positive definite matrix square root is `sqrtm(A)`. This is different from `sqrt(A)`, which is the *element-wise* square root (not the same!).

In the single-variable case, we computed the probability that $x \sim \mathcal{N}(\mu, \sigma)$ lies in some symmetric interval about the mean, i.e., $\mathbf{P}(\mu - a \leq x \leq \mu + a)$. The analogous quantity for a multivariate Gaussian is the *confidence ellipsoid*.

Since the density $f(x)$ is proportional to $\exp\left(-\frac{1}{2}(x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu)\right)$, the contours of constant density (and their interiors) are given by the nested sets:

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

This set is an ellipsoid. To see why, make the standard normal substitution:

$$S_\alpha = \left\{ x \in \mathbb{R}^n \;\middle|\; \left\|\Sigma^{-1/2}(x-\mu)\right\|^2 \leq \alpha \right\} = \left\{ x \in \mathbb{R}^n \;\middle|\; \|z\|^2 \leq \alpha \right\}$$

We have seen this sort of ellipse before. It is a standard control ellipsoid:

$$\left\{ x \;\middle|\; \left\|\Sigma^{-1/2}(x-\mu)\right\|^2 \leq \alpha \right\} = \left\{ \mu + \Sigma^{1/2} z \;\middle|\; \|z\| \leq \alpha \right\} = \mu + \left\{ \sqrt{\alpha}\Sigma^{1/2} w \;\middle|\; \|w\| \leq 1 \right\}$$

To plot this, we should compute the SVD (or eigenvalue) decomposition as in (6). Then the confidence ellipsoid is centered at μ and has its axes pointing in the u_i directions, with corresponding lengths $\sqrt{\alpha}\sqrt{\lambda_i}$.

The probability associated with this level ellipsoid is:

$$\begin{aligned} p &= \mathbf{P}(x \in S_\alpha,\; x \sim \mathcal{N}(\mu, \Sigma)) \\ &= \mathbf{P}(\|z\|^2 \leq \alpha,\; z \sim \mathcal{N}(0, I)) \\ &= \mathbf{P}(z_1^2 + \cdots + z_n^2 \leq \alpha,\; z_i \sim \mathcal{N}(0,1)) \\ &= \mathbf{P}(w \leq \alpha,\; w \sim \chi_n^2) \\ &= F_{\chi_n^2}(\alpha) \end{aligned}$$

In other words, the CDF of the Chi-squared distribution! To find which α corresponds to a desired probability p, we can invert this cdf:

$$\alpha = F_{\chi_n^2}^{-1}(p)$$

Some useful MATLAB commands for computing the CDF and inverse CDF of a Chi-squared distribution:

- CDF: `p = chi2cdf(alpha,n)`
- Inverse CDF: `alpha = chi2inv(p,n)`

Fig. 5 below shows an example of a $n = 2$ dimensional Gaussian with various contours shown.

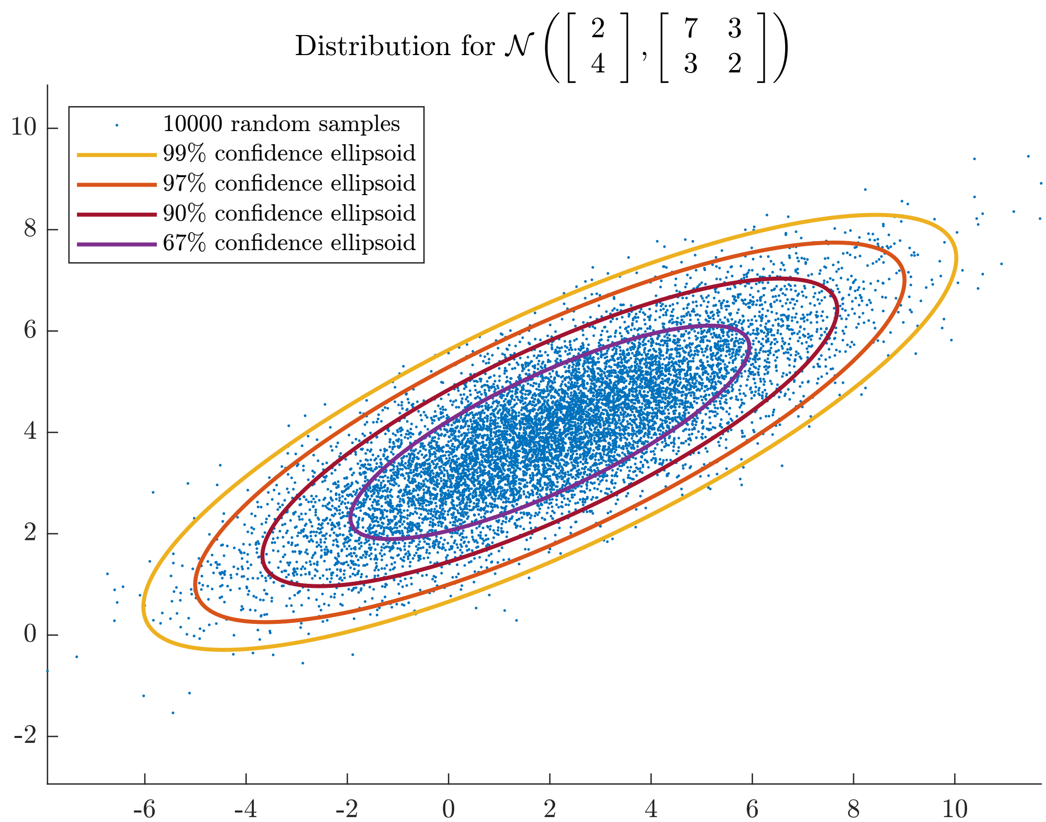

Figure 5: Confidence ellipsoids for different values of p. Each ellipse contains on average a fraction p of all the points. If $x \sim \mathcal{N}(\mu, \Sigma)$, then the confidence ellipsoid for a given p is the ellipsoid $\left\{x \in \mathbb{R}^n \;\middle|\; (x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu) \leq \alpha\right\}$, where $\alpha = F^{-1}_{\chi^2_n}(p)$. The ellipsoid is centered at μ and its axes point in the directions u_i with lengths $\sqrt{\alpha}\sqrt{\lambda_i}$, where (λ_i, u_i) are the eigenvalue-eigenvector pairs of the covariance matrix Σ.