RBOT101: MATHEMATICAL FOUNDATIONS OF ROBOTICS

SUMMER 2021

Expectation and Moments

Benjamin Gravell

The University of Texas at Dallas

Department of Mechanical Engineering

P3-1.

We are considering a Gaussian random variable $X \sim \mathcal{N}(\mu, \sigma^2)$.

From the definition of expectation, we directly evaluate the variance of X

$$\begin{aligned}
\mathbb{E}[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx &\quad (1)\\
&= \int_{-\infty}^{\infty} z^2 \frac{\sigma^2}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz &\quad (2)\\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) dz &\quad (3)
\end{aligned}$$

where the last step is by using the change of variables $z = (x-\mu)/\sigma$.

Next, integrate by parts with $u = z$ and $dv = z\exp(-z^2/2)$ yielding $du = dz$ and $v = -\exp(-z^2/2)$, so that the integral becomes

$$\begin{aligned}
\int_{-\infty}^{\infty} z^2 \exp(-z^2/2)dz &= \left(-z\exp(-z^2/2)\right)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp(-z^2/2)dz &\quad (4)\\
&= -0 + 0 + \sqrt{2\pi}, &\quad (5)
\end{aligned}$$

where the last term is because the standard normal $\mathcal{N}(0,1)$ pdf integrates to 1 over the real line.

Thus we showed

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2 \quad (6)$$

P3-2.

Recall the exponential distribution has the cdf

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The exact tail probability for $\lambda = 1$ and $t = 2$ is therefore

$$P[X \geq 2] = 1 - P[X \leq 2] = 1 - F_X(2) = 1 - (1 - \exp(-\lambda x)) = \exp(-2) \approx 0.136. \tag{8}$$

Using the Markov inequality, we obtain the tail bound

$$P[X \geq 2] \leq \frac{\mathbb{E}[X]}{2} = \frac{1}{2\lambda} = 0.5. \tag{9}$$

In this instance, the Markov inequality is rather loose, only guaranteeing X smaller than 2 with probability 50% compared with the true probability of 86.4%.

P3-3.

Using either the `cdf()` method of `scipy.stats.norm` or a table of the standard normal cdf, we obtain the exact concentration probability

$$\begin{aligned}
\mathbb{P}[|X-\mu| \geq t] &= \mathbb{P}[X - mu \geq t \cup X - mu \leq -t] && \text{(unfold abs. value)} \\
&= \mathbb{P}[X - mu \geq t] + \mathbb{P}[X - mu \leq -t] && \text{(probability of disjoint union)} \\
&= \mathbb{P}[X \geq 2] + \mathbb{P}[X \leq -2] && (10) \\
&= 1 - \mathbb{P}[X \leq 2] + \mathbb{P}[X \leq -2] && (11) \\
&\approx 1 - 0.9772 + 1 - 0.9772 && (12) \\
&= 0.0456 && (13)
\end{aligned}$$

Using the Chebyshev inequality we have

$$\begin{aligned}
\mathbb{P}[|X| \geq 2] &= \mathbb{P}[|X-\mu| \geq t] && (14) \\
&\leq \frac{\sigma^2}{t^2} = \frac{1}{2^2} = 0.25 && (15)
\end{aligned}$$

In this instance, the Chebyshev inequality is rather loose, only guaranteeing $|X-\mu|$ smaller than 2 with probability 75% compared with the true probability of 95.4%.

P3-4.

We have

$$
\begin{aligned}
c_{11} &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] && \text{(by definition of } c_{11}\text{)} \\
&= \mathbb{E}[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y])] && (16) \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]) + \mathbb{E}[X]\mathbb{E}[Y] && \text{(linearity of } \mathbb{E}[\cdot]\text{)} \\
&= m_{11} - \mathbb{E}[X]\mathbb{E}[Y] && \text{(by definition of } m_{11}\text{)}
\end{aligned}
$$

and rearranging yields the claim.

Remark: This is essentially the same as the relation between second moment and variance of a random variable ("auto-"), but replacing the second instance of X with Y ("cross-").

P3-5.

Since we proved

$$c_{11} = m_{11} - \mathbb{E}[X]\mathbb{E}[Y] \tag{17}$$

it is immediately clear that

$$c_{11} = m_{11} \tag{18}$$

if and only if $\mathbb{E}[X]\mathbb{E}[Y] = 0$. That is, the correlation and the covariance are equal only when at least one of the random variables has zero mean.

P3-6.

Consider the expression

$$\mathbb{E}[[\lambda(X-\mathbb{E}[X])-(Y-\mathbb{E}[Y])]^2] \tag{19}$$

where λ is any real constant. We will show this expression is nonnegative by writing the expectation in integral form and treating λ as a parameter:

$$Q(\lambda)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\lambda(X-\mathbb{E}[X])-(Y-\mathbb{E}[Y])]^2 f_{XY}(x,y)dxdy \geq 0 \tag{20}$$

where the inequality follows because the integrand is always square and thus everywhere nonnegative. Expanding the square and evaluating expectations of each term yields

$$Q(\lambda)=\lambda^2 c_{20}+c_{02}-2\lambda c_{11} \geq 0, \tag{21}$$

which is a nonnegative quadratic in λ. Therefore $Q(\lambda)$ has at most a single real root, and therefore its discriminant must satisfy

$$\left(\frac{c_{11}}{c_{20}}\right)^2-\frac{c_{02}}{c_{20}} \leq 0 \tag{22}$$

which can be rearranged to

$$\frac{c_{11}^2}{c_{02}c_{20}} \leq 1 \tag{23}$$

i.e.

$$\rho^2 \leq 1 \quad \leftrightarrow \quad |\rho| \leq 1 \tag{24}$$

P3-7.

We have

$$
\begin{aligned}
\sigma^2_{X+Y} &= \mathbb{E}\left[((X+Y)-\mathbb{E}[X+Y]))^2\right] && \text{(definition of } \sigma^2\text{)} \\
&= \mathbb{E}\left[((X-\mathbb{E}[X])+(Y-\mathbb{E}[Y]))^2\right] && \text{(linearity of } \mathbb{E}[\cdot]\text{)} \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2+2(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])+(Y-\mathbb{E}[Y])^2\right] && (25) \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right]+2\mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]+\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] && \text{(linearity of } \mathbb{E}[\cdot]\text{)} \\
&= \mathbb{E}\left[(X-\mathbb{E}[X])^2\right]+2\cdot 0+\mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] && \text{(assumption of } X, Y \text{ uncorrelated)} \\
&= \sigma^2_X+\sigma^2_Y && \text{(definition of } \sigma^2\text{)}
\end{aligned}
$$

P3-8.

We have

$$
\begin{aligned}
\mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x,y)\,dx\,dy && (26) \\
&= \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y f_Y(y)\,dy\right) && \text{(by independence assumption)} \\
&= \mathbb{E}[X]\mathbb{E}[Y] && (27)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y] && \text{(proved earlier)} \\
&= 0 && \text{(by (27))}
\end{aligned}
$$

i.e. X and Y are uncorrelated.

P3-9.

In an earlier problem we established that the correlation and the covariance are equal only when at least one of the random variables has zero mean.

For X and Y to be both uncorrelated and orthogonal, the correlation and covariance must both be equal to zero, which only happens when at least one of the random variables has zero mean.

P3-10.

We have

$$
\begin{aligned}
K &= \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^\intercal] && \text{(by definition of } K) \\
&= \mathbb{E}[XX^\intercal - \mathbb{E}[X]X^\intercal - X\mathbb{E}[X]^\intercal + \mathbb{E}[X]\mathbb{E}[X]^\intercal)] && (28) \\
&= \mathbb{E}[XX^\intercal] - \mathbb{E}[X]\mathbb{E}[X]^\intercal - \mathbb{E}[X]\mathbb{E}[X]^\intercal + \mathbb{E}[X]\mathbb{E}[X]^\intercal && \text{(linearity of } \mathbb{E}[\cdot]) \\
&= R - \mu\mu^\intercal && \text{(by definition of } R,\ \mu)
\end{aligned}
$$

and rearranging yields the claim.

Remark: This is essentially the same as the relation between second moment and variance of a scalar random variable, but now posed in the multivariate setting.

P3-11.

We have

$$
\begin{aligned}
D &= \mathbb{E}\left[\begin{bmatrix} X-\mu_X \\ Y-\mu_Y \end{bmatrix}\begin{bmatrix} X-\mu_X \\ Y-\mu_Y \end{bmatrix}^\intercal\right] && \text{(by definition of } D) \\
&= B - \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}^\intercal && (29)
\end{aligned}
$$

where the last line follows by the second moment relation $K = R - \mu\mu^\intercal$ and considering the stacked random vector $Z = [X^\intercal \ Y^\intercal]^\intercal$.

P3-12.

Consider R and K related to X. First we show that $K \succeq 0$ by using the equivalent condition $w^\intercal K w \geq 0$ for any constant vector w. We have

$$
\begin{aligned}
w^\intercal K w &= w^\intercal \mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^\intercal] w && \text{(definition of } K\text{)} \\
&= \mathbb{E}[w^\intercal (X-\mathbb{E}[X])(X-\mathbb{E}[X])^\intercal w] && \text{(linearity of } \mathbb{E}[\cdot]\text{)} \\
&= \mathbb{E}[[w^\intercal (X-\mathbb{E}[X])]^2] && (w^\intercal (X-\mathbb{E}[X]) \text{ is a scalar}) \\
&= \mathbb{E}[V^2] && \text{(define random variable } V = w^\intercal (X-\mathbb{E}[X])\text{)} \\
&= \int_{-\infty}^{\infty} v^2 f_V(v) dv && \text{(definition of } \mathbb{E}[\cdot]\text{)} \\
&\geq 0 && (30)
\end{aligned}
$$

where the final inequality follows because the integrand is everywhere nonnegative, specifically $v^2 \geq 0$ for any v and $f_V(v) \geq 0$ for any v by properties of pdfs. This shows that indeed $K \succeq 0$.

Next, $R \succeq K$ follows immediately by the property

$$R = K + \mu\mu^\intercal \tag{31}$$

and

$$\mu\mu^\intercal \succeq 0. \tag{32}$$

The relation for B and D related to both X and Y follows immediately by considering the stacked random vector $Z = [X^\intercal \ Y^\intercal]^\intercal$.

Remark: If you have not proved the general property

$$AA^\intercal \succeq 0 \tag{33}$$

for any matrix A, it is easy to prove. Just consider an arbitrary vector x and the vector $y = A^\intercal x$, then

$$x^\intercal A A^\intercal x = y^\intercal y \geq 0. \tag{34}$$

P3-13.

The multivariate Gaussian pdf is

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left[-\frac{1}{2}(x-\mu)^\intercal K^{-1}(x-\mu)\right] \tag{35}$$

Since X is assumed white, it has zero mean and identity covariance, so the pdf becomes

$$\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{(2\pi)^n \det(I)}} \exp\left[-\frac{1}{2}(x-0)^\intercal I^{-1}(x-0)\right] && (36)\\
&= \frac{1}{\sqrt{(2\pi)^n}} \exp\left[-\frac{1}{2}x^\intercal x\right] && (37)\\
&= \frac{1}{\sqrt{(2\pi)^n}} \exp\left[-\frac{1}{2}\sum_{i=1}^n x_i^2\right] && (38)\\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x_i^2\right]. && (39)
\end{aligned}$$

Next we evaluate the marginal pdf for the ith component:

$$\begin{aligned}
f_{X_i}(x_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n && (40)\\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x_j^2\right] dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n && (41)\\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x_i^2\right] \times \prod_{j\neq i} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x_j^2\right] dx_j && (42)\\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x_i^2\right], && (43)
\end{aligned}$$

where the last line follows by recognizing each term in the product over $i \neq j$ as the integral of a Gaussian pdf, which evaluates to unity. Thus, the marginals are simply standard normal pdfs.

Also, by inspection of the preceding development it is clear that the pdf $f_X(x)$ of X is the product of the marginal pdfs $f_{X_i}(x_i)$ of each component X_i, from which we conclude that indeed the components of X are statistically independent.

Remark: A similar calculation shows that the components are independent when $\mu = 0$ and K is diagonal.