Limit theorems

Limit theorems for sums (1/2)

Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed (i.i.d.) random variables (square integrable). The empirical mean is the random variable:

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

Its mean is $\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_1]$, and its variance is

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

with $\sigma^2 = \mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n)$. This quantity goes to 0 as $n \to \infty$, which means that $\bar{X}_n$ concentrates on the deterministic value $\mathbb{E}[X_1]$:

$$\mathbb{P}\left(|\bar{X}_n - \mathbb{E}[X_1]| \geq \varepsilon\right) \leq \frac{\mathbb{E}[|\bar{X}_n - \mathbb{E}[X_1]|^2]}{\varepsilon^2} = \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon}$$

- (Strong) law of large numbers.

Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables (square integrable). The empirical mean $\bar{X}_n$ converges to $\mathbb{E}[X_1]$ with probability one:

$$\bar{X}_n \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1] \quad a.s. \qquad \text{or } \mathbb{P}\left(\lim_{n\to\infty}\bar{X}_n = \mathbb{E}[X_1]\right) = 1$$

Limit theorems for sums (2/2)

- Central limit theorem

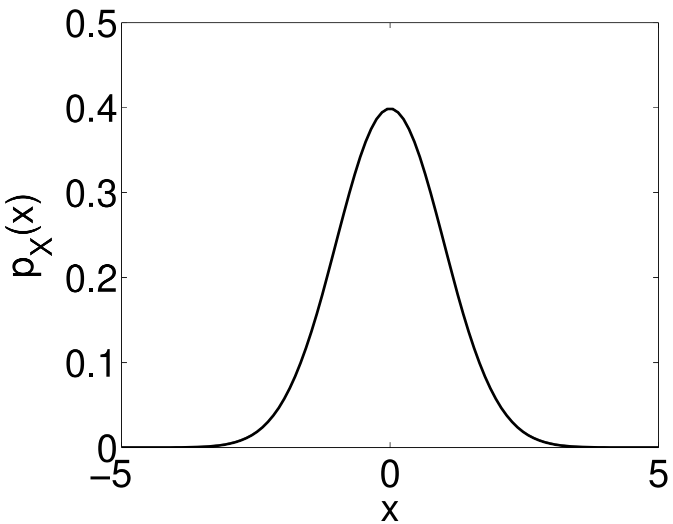
Let $(X_n)_{n\geq 0}$ be a sequence of independent and identically distributed random variables (square integrable), with the mean $\mu \in \mathbb{R}$ and the variance σ^2, $\sigma \in (0,\infty)$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma^2)$$

This means that, for any interval $I \subset \mathbb{R}$,

$$\mathbb{P}\left(\sqrt{n}(\bar{X}_n - \mu) \in I\right) \overset{n\to\infty}{\longrightarrow} \int_I p_{0,\sigma^2}(x)dx$$

where $p_{0,\sigma^2}(x) = (2\pi\sigma^2)^{-1/2}\exp(-x^2/(2\sigma^2))$.



Limit theorems for maxima (1/2)

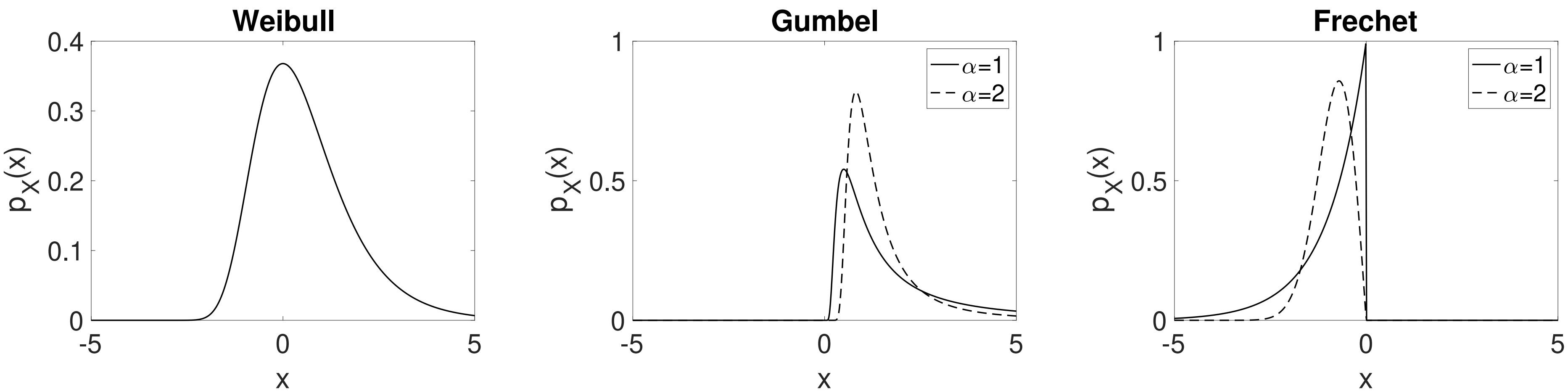
Let $M_n = \max(X_1, \ldots, X_n)$ where the X_i are i.i.d. with pdf $p(x)$.

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = \mathbb{P}(X \leq x)^n = \left(1 - \int_x^\infty p(s)ds\right)^n$$

Theorem (Fisher-Tippett-Gnedenko): Assume that there exist a_n and b_n such that $\mathbb{P}(M_n \leq a_n + b_n x)$ converges to a cdf $F(x)$. Then $F(x)$ belongs to one of the three following types (up to affine scaling):

- $F_1(x) = e^{-e^{-x}}$ with support $\mathbb{R}$ (Weibull).
- $F_{2,\alpha}(x) = e^{-1/x^\alpha}$ with support $\mathbb{R}^+$, $\alpha > 0$ (Gumbel).
- $F_{3,\alpha}(x) = e^{-(-x)^\alpha}$ with support $\mathbb{R}^-$, $\alpha > 0$ (Fréchet).

Conclusion: As $n \to \infty$, $M_n = a_n + b_n Y$ where Y follows one of the three "extreme" distributions.



Limit theorems for maxima (2/2)

Main idea to determine the asymptotic distribution of $M_n = \max(X_1, \ldots, X_n)$:
1) Denote $x_{n,\alpha}$ such that $\int_{x_{n,\alpha}}^{\infty} p(s)ds = \frac{\alpha}{n}$. Then $\mathbb{P}(M_n \leq x_{n,\alpha}) = (1 - \frac{\alpha}{n}) \to e^{-\alpha}$.
2) Look for the expansion of $x_{n,\alpha}$ as $n \to \infty$.

Example 1: $X_i \sim \mathcal{U}(0,1)$, i.e. $p(x) = \mathbf{1}_{[0,1]}(x)$.
Then $a_n = 1$, $b_n = 1/n$, and $F = F_{3,1}$.

Example 2: $X_i \sim \mathcal{E}(1)$, i.e. $p(x) = e^{-x}\mathbf{1}_{[0,\infty)}(x)$.
Then $a_n = \ln(n)$, $b_n = 1$, and $F = F_1$.

Example 3: $X_i \sim \mathcal{N}(0,1)$, i.e. $p(x) = (2\pi)^{-1/2} e^{-x^2/2}$.
Then $a_n = \sqrt{2\ln(n)} - \ln\ln n/\sqrt{8\ln(n)} - \ln(4\pi)/\sqrt{8\ln(n)}$, $b_n = 1/\sqrt{2\ln(n)}$, and $F = F_1$.

Ergodic theory

Let $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ be a stationary random process with mean $\mu = \mathbb{E}[Z(\mathbf{0})] = \mathbb{E}[Z(\boldsymbol{x})]$.

Ergodic Theorem. If Z satisfies the ergodic hypothesis, then

$$\frac{1}{N^d}\int_{[0,N]^d} Z(\boldsymbol{x})d\boldsymbol{x} \stackrel{N\to\infty}{\longrightarrow} \mu \quad \text{with probability 1}$$

Ergodic hypothesis = "the orbit $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ visits all of phase space".

Ergodic theorem = "the spatial average is equivalent to the statistical average".

Counter-example for the ergodic hypothesis:
Let Z_1 and Z_2 be stationary, both satisfy the ergodic theorem, $\mu_j = \mathbb{E}[Z_j(\boldsymbol{x})]$, $j = 1, 2$, with $\mu_1 \neq \mu_2$.
Let χ be a random variable $\mathbb{P}(\chi = 0) = \mathbb{P}(\chi = 1) = 1/2$ (independently of Z_j).
Let $Z(\boldsymbol{x}) = \chi Z_1(\boldsymbol{x}) + (1-\chi)Z_2(\boldsymbol{x})$.
Z is a stationary process with mean $\mu = \frac{1}{2}(\mu_1 + \mu_2)$.

$$\begin{aligned}\frac{1}{N^d}\int_{[0,N]^d} Z(\boldsymbol{x})d\boldsymbol{x} &= \chi\left(\frac{1}{N^d}\int_{[0,N]^d} Z_1(\boldsymbol{x})d\boldsymbol{x}\right) + (1-\chi)\left(\frac{1}{N^d}\int_{[0,N]^d} Z_2(\boldsymbol{x})d\boldsymbol{x}\right)\\ &\stackrel{N\to\infty}{\longrightarrow} \chi\mu_1 + (1-\chi)\mu_2\end{aligned}$$

which is a random limit different from μ.
The limit depends on χ because Z has been trapped in a part of phase space.

Mean square theory

Let $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ be a stationary random process with mean μ and covariance

$$c(\boldsymbol{x}) = \mathbb{E}\left[(Z(\boldsymbol{y})-\mu)(Z(\boldsymbol{y}+\boldsymbol{x})-\mu)\right].$$

By stationarity, c is an even function:

$$\begin{aligned} c(-\boldsymbol{x}) &= \mathbb{E}\left[(Z(\boldsymbol{y})-\mu)(Z(\boldsymbol{y}-\boldsymbol{x})-\mu)\right] = \mathbb{E}\left[(Z(\boldsymbol{y}'+\boldsymbol{x})-\mu)(Z(\boldsymbol{y}')-\mu)\right] \\ &= c(\boldsymbol{x}). \end{aligned}$$

By Cauchy-Schwarz inequality, c reaches its maximum at 0:

$$c(\boldsymbol{x}) \leq \mathbb{E}\left[(Z(\boldsymbol{y})-\mu)^2\right]^{1/2}\mathbb{E}\left[(Z(\boldsymbol{y}+\boldsymbol{x})-\mu)^2\right]^{1/2} = c(\mathbf{0}),$$

and $c(\mathbf{0}) = \mathrm{Var}(Z(\mathbf{0}))$.

Assume that $\int_{\mathbb{R}^d}|c(\boldsymbol{x})|d\boldsymbol{x} < \infty$. Let

$$S(N) = \frac{1}{N^d}\int_{[0,N]^d} Z(\boldsymbol{x})d\boldsymbol{x}.$$

Then

$$\mathbb{E}\left[(S(N)-\mu)^2\right] \stackrel{N\to\infty}{\longrightarrow} 0,$$

more exactly

$$N\mathbb{E}\left[(S(N)-\mu)^2\right] \stackrel{N\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} c(\boldsymbol{x})d\boldsymbol{x}.$$

Proof when $d = 1$:

$$\begin{aligned}
\mathbb{E}\left[(S(N)-\mu)^2\right] &= \mathbb{E}\left[\frac{1}{N^2}\int_0^N dt_1 \int_0^N dt_2 (Z(t_1)-\mu)(Z(t_2)-\mu)\right] \\
&\overset{\text{symmetry}}{=} \frac{2}{N^2}\int_0^N dt_1 \int_0^{t_1} dt_2 c(t_1 - t_2) \\
&\overset{\substack{\tau = t_1 - t_2 \\ h = t_2}}{=} \frac{2}{N^2}\int_0^N d\tau \int_0^{N-\tau} dh c(\tau) \\
&= \frac{2}{N^2}\int_0^N d\tau (N-\tau) c(\tau) = \frac{2}{N}\int_0^\infty d\tau c_N(\tau),
\end{aligned}$$

where $c_N(\tau) = c(\tau)(1-\tau/N)\mathbf{1}_{[0,N]}(\tau)$. By Lebesgue's convergence theorem:

$$N\mathbb{E}\left[(S(N)-\mu)^2\right] \overset{N\to\infty}{\longrightarrow} 2\int_0^\infty c(\tau)d\tau,$$

Note that the $L^2(\mathbb{P})$ convergence implies convergence in probability as the limit is deterministic. Indeed, by Chebychev inequality, for any $\delta > 0$,

$$\mathbb{P}\left(|S(N)-\mu| \geq \delta\right) \leq \frac{\mathbb{E}\left[(S(N)-\mu)^2\right]}{\delta^2} \overset{N\to\infty}{\longrightarrow} 0.$$

Note also that we can obtain by the same method that, for any $\boldsymbol{k} \in \mathbb{R}^d$,

$$N^d \mathbb{E}\left[\Big| \int_{[0,N]^d} \big(Z(\boldsymbol{x}) - \mu\big) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} d\boldsymbol{x} \Big|^2 \right] \overset{N\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} c(\boldsymbol{x}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}} d\boldsymbol{x},$$

which shows that the Fourier transform of the covariance function of a stationary process is nonnegative. This is a preliminary form of Bochner's theorem: a function $c(\boldsymbol{x})$ is a covariance function of a stationary process if and only if its Fourier transform is nonnegative.

Extrema of Gaussian processes

Local extrema of a Gaussian process

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x} \in \mathbb{R}^d}$ with mean zero and covariance function $C(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$.

The local form of a Gaussian process around a local extremum at $\boldsymbol{x}_0$ with peak value $z_0 \gg 1$ is essentially deterministic and given by the covariance function:

$$Z(\boldsymbol{x}) = z_0 c(\boldsymbol{x} - \boldsymbol{x}_0) + O(1)$$

Local extrema of a Gaussian process

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ with mean zero and covariance function $C(\boldsymbol{x})$, such that $c(\mathbf{0})=1$.

The local form of a Gaussian process around a local extremum at $\boldsymbol{x}_0$ with peak value $z_0 \gg 1$ is essentially deterministic and given by the covariance function:

$$Z(\boldsymbol{x}) = z_0 c(\boldsymbol{x}-\boldsymbol{x}_0) + O(1)$$

Proof: Gaussian conditioning.

We have

$$\begin{pmatrix} Z(\boldsymbol{x}) \\ Z(\boldsymbol{x}') \\ Z(\boldsymbol{x}_0) \\ \nabla Z(\boldsymbol{x}_0) \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \\ 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} 1 & c(\boldsymbol{x}-\boldsymbol{x}') & c(\boldsymbol{x}-\boldsymbol{x}_0) & -\nabla c(\boldsymbol{x}-\boldsymbol{x}_0)^T \\ c(\boldsymbol{x}-\boldsymbol{x}') & 1 & c(\boldsymbol{x}'-\boldsymbol{x}_0) & -\nabla c(\boldsymbol{x}'-\boldsymbol{x}_0)^T \\ c(\boldsymbol{x}-\boldsymbol{x}_0) & c(\boldsymbol{x}'-\boldsymbol{x}_0) & 1 & \mathbf{0}^T \\ -\nabla c(\boldsymbol{x}-\boldsymbol{x}_0) & -\nabla c(\boldsymbol{x}'-\boldsymbol{x}_0) & \mathbf{0} & \mathbf{H} \end{pmatrix} \Big)$$

where $\mathbf{H} = \mathbb{E}[\nabla Z(\mathbf{0}) \nabla Z(\mathbf{0})^T] = -\big(\partial^2_{x_j x_l} c(\mathbf{0})\big)_{j,l=1}^d$.

The distribution of $(Z(\boldsymbol{x}), Z(\boldsymbol{x}'))^T$ given $Z(\boldsymbol{x}_0) = z_0$ and $\nabla Z(\boldsymbol{x}_0) = \mathbf{0}$ is

$$\mathcal{L}\Big(\begin{pmatrix} Z(\boldsymbol{x}) \\ Z(\boldsymbol{x}') \end{pmatrix} \Big| \begin{pmatrix} Z(\boldsymbol{x}_0) \\ \nabla Z(\boldsymbol{x}_0) \end{pmatrix} = \begin{pmatrix} z_0 \\ \mathbf{0} \end{pmatrix} \Big) = \mathcal{N}\Big(\begin{pmatrix} \mu_{\mathrm{p}}(\boldsymbol{x}) \\ \mu_{\mathrm{p}}(\boldsymbol{x}') \end{pmatrix}, \begin{pmatrix} \sigma_{\mathrm{p}}^2(\boldsymbol{x}) & c_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{x}') \\ c_{\mathrm{p}}(\boldsymbol{x}, \boldsymbol{x}') & \sigma_{\mathrm{p}}^2(\boldsymbol{x}') \end{pmatrix} \Big)$$

with

$$\begin{aligned} \mu_{\mathrm{p}}(\boldsymbol{x}) &= \begin{pmatrix} c(\boldsymbol{x} - \boldsymbol{x}_0) & -\nabla c(\boldsymbol{x} - \boldsymbol{x}_0)^T \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{H} \end{pmatrix}^{-1} \begin{pmatrix} z_0 \\ \mathbf{0} \end{pmatrix} \\ &= c(\boldsymbol{x} - \boldsymbol{x}_0) z_0 \end{aligned}$$

and

$$\begin{aligned} \sigma_{\mathrm{p}}^2(\boldsymbol{x}) &= 1 - \begin{pmatrix} c(\boldsymbol{x} - \boldsymbol{x}_0) & -\nabla c(\boldsymbol{x} - \boldsymbol{x}_0)^T \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{H} \end{pmatrix}^{-1} \begin{pmatrix} c(\boldsymbol{x} - \boldsymbol{x}_0) \\ -\nabla c(\boldsymbol{x} - \boldsymbol{x}_0) \end{pmatrix} \\ &= 1 - c(\boldsymbol{x} - \boldsymbol{x}_0)^2 - \nabla c(\boldsymbol{x} - \boldsymbol{x}_0)^T \mathbf{H}^{-1} \nabla c(\boldsymbol{x} - \boldsymbol{x}_0) \end{aligned}$$

Number of local maxima of a Gaussian process

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ with mean zero and covariance function $c(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$.

The mean number of local maxima of a Gaussian process in a domain Ω is determined by the second and fourth-order derivatives of c at $\mathbf{0}$.

$\hookrightarrow$ Rice formula.

Number of local maxima of a Gaussian process

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ with mean zero and covariance function $c(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$.

The mean number of local maxima of a Gaussian process in a domain Ω is determined by the second and fourth-order derivatives of c at $\mathbf{0}$.

Rice formula: Step 1: If $f(x)$ is a smooth function from $\mathbb{R}$ to $\mathbb{R}$, non-degenerate (i.e. $f'' \neq 0$ at any extremal point where $f' = 0$), then the number of extrema of f over an open interval Ω is

$$N^{\Omega} = \int_{\Omega} \delta(f'(x))|f''(x)|dx$$

Step 2: If $f(\boldsymbol{x})$ is a smooth function, non-degenerate (i.e. the Hessian f'' of f is not singualr at any extremum where the gradient $f' = \mathbf{0}$), then the number of extrema of f over Ω is

$$N^{\Omega} = \int_{\Omega} \delta(f'(\boldsymbol{x}))|\det f''(\boldsymbol{x})|d\boldsymbol{x}$$

Step 3: If $f(\boldsymbol{x})$ is a smooth function, non-degenerate, then the number of maxima of f over Ω whose values are larger than u is

$$N_u^{\Omega} = \int_{\Omega} \delta(f'(\boldsymbol{x}))\mathbf{1}_{f(\boldsymbol{x})\geq u, f''(\boldsymbol{x})<0}|\det f''(\boldsymbol{x})|d\boldsymbol{x}$$

Step 4: If $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ is a stationary Gaussian process with mean zero and smooth covariance function $c(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$, then the mean number of maxima of Z over Ω whose values are larger than u is

$$\mathbb{E}[N_u^\Omega] = |\Omega| \int dy \int d\mathbf{y}'' p_{Z(\mathbf{0}),Z'(\mathbf{0}),Z''(\mathbf{0})}(y, \mathbf{0}, \mathbf{y}'')| \det \mathbf{y}''|\mathbf{1}_{y\geq u,\mathbf{y}''<0}$$

where $p(y, \boldsymbol{y}', \mathbf{y}'')$ is the probability density function of $(Z(\mathbf{0}), Z'(\mathbf{0}), Z''(\mathbf{0}))$.

In dimension $d = 1$, we have $\begin{pmatrix} Z(0) \\ Z'(0) \\ Z''(0) \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -H \\ 0 & H & 0 \\ -H & 0 & K \end{pmatrix} \Big)$ where $H = \mathbb{E}[Z'(0)^2] = -c''(0)$ and $K = \mathbb{E}[Z''(0)^2] = c^{(4)}(0)$.

This shows that $(Z(0), Z''(0))$ and $Z'(0)$ are independent:

$$p_{Z(0),Z'(0),Z''(0)}(y, y', y'') = p_{Z'(0)}(y')p_{Z(0),Z''(0)}(y, y''),$$

$$p_{Z(0),Z'(0),Z''(0)}(y, 0, y'') = \frac{1}{\sqrt{2\pi H}} p_{Z(0),Z''(0)}(y, y'')$$

and that the distribution of $Z''(0)$ given $Z(0) = y$ is $\mathcal{N}(-Hy, K - H^2)$:

$$\int dy'' p_{Z''(0)|Z(0)}(y''|y)|y''|\mathbf{1}_{y''<0} \simeq Hy \text{ for } y \gg 1$$

Therefore $\mathbb{E}[N_u^\Omega] \simeq |\Omega| \frac{\sqrt{H}}{2\pi} \int_u^\infty y \exp\big(-\frac{y^2}{2}\big) dy = |\Omega| \frac{\sqrt{H}}{2\pi} \exp\big(-\frac{u^2}{2}\big)$

Let us consider a stationary Gaussian process $(Z(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^d}$ with mean zero and covariance function $c(\boldsymbol{x})$, such that $c(\mathbf{0}) = 1$.

The number of local maxima of a Gaussian process in a domain Ω is essentially deterministic (when the volume $|\Omega|$ is larger than the hotspot volume) and given by the number of hotspot volumes that fits in Ω.

The hotspot volume V_c is the typical volume occupied by a local maximum. From the description of a local maximum, it is the essential support of the covariance function:

$$V_c = (2\pi)^{(d+1)/2}(\det \mathbf{H})^{-1/2} \quad \text{with } \mathbf{H} = \big(- \partial_{x_j}\partial_{x_l} c)(\mathbf{0})\big)_{j,l=1,\ldots,d}$$

Cf. R. Adler's book: When $|\Omega| \gg V_c$, the number N_u^Ω of local maxima of $Z(\boldsymbol{x})$ in a domain Ω that have peak values larger than u is

$$N_u^\Omega \simeq \frac{|\Omega|}{V_c} u^{d-1} e^{-u^2/2}$$