PROBLEM 1

Given the joint CDF of random variables X_1 and X_2,

$$F_{X_1,X_2}(x_1,x_2) = 1 - \exp(-x_1) - \exp(-x_2) + \exp(-x_1 - x_2 - x_1 x_2), \quad x_1, x_2 \geq 0, \tag{1}$$

we are tasked with finding the marginal CDF $F_{X_1}(x_1)$ and the conditional CDF $F_{X_2|X_1}(x_2|x_1)$, and subsequently generating realizations of X_1, X_2 using the inversion method.

The marginal CDF of X_1 is trivially calculated in the limit $x_2 \to \infty$ as

$$\boxed{F_{X_1}(x_1)} = F_{X_1,X_2}(x_1,\infty) = 1 - \exp(-x_1). \tag{2}$$

Applying the relation

$$F_{X_2|X_1}(x_2|x_1) = \left(\int_0^{x_2} f_{X_1,X_2}(x_1,t_2) dt_2 \right) \Big/ f_{X_1}(x_1), \tag{3}$$

where the marginal and joint pdfs are

$$f_{X_1}(x_1) = \partial_{x_1} F_{X_1}(x_1)$$
$$f_{X_1,X_2}(x_1,x_2) = \partial_{x_1} \partial_{x_2} F_{X_1,X_2}(x_1,x_2),$$

it can be shown that

$$\boxed{F_{X_2|X_1}(x_2|x_1)} = 1 - (1 + x_2)\exp(-[1 + x_1]x_2), \tag{4}$$

which is impossible to invert analytically, though computational root-finding methods show success.

Realizations of X_1, X_2 are generated in the standard manner: for each $i = 1, ..., N$, a random variable $U_1^i \sim U[0,1]$ is generated, and set equal to $F_{X_1}(x_1)$, which can be solved for realization x_1^i. Another random variable $U_2^i \sim U[0,1]$ is generated and set equal to $F_{X_2|X_1}(x_2|x_1)$, which is then solved numerically for x_2^i, given x_1^i. We choose Matlab's `fzero` function as our root finder.

In Figure 1, we generate $N = 10{,}000$ realizations and compare the cumulative expectation of the first n samples to the analytical expectations

$$\begin{aligned} \langle x_1 \rangle &= 1.0 \\ \langle x_1 \rangle &= 1.0 \\ \langle x_1 x_2 \rangle &= 0.596347 \end{aligned} \tag{5}$$

All three quantities approach their analytical values, and the relative error in each quantity is seen to decrease as n increases.

This method of verification is by no means rigorous. The mean square error of the empirical CDF or pdf would be a better way of checking the validity of our answers, but this suffices for the purposes of this exercise.

PROBLEM 2

This problem concerns a derivation from scratch of the Bayesian MAP estimate of a random variable V, assuming a Gaussian prior $V \sim N(V_0, \sigma_0^2)$. Further details are worked by hand on the attached sheets.

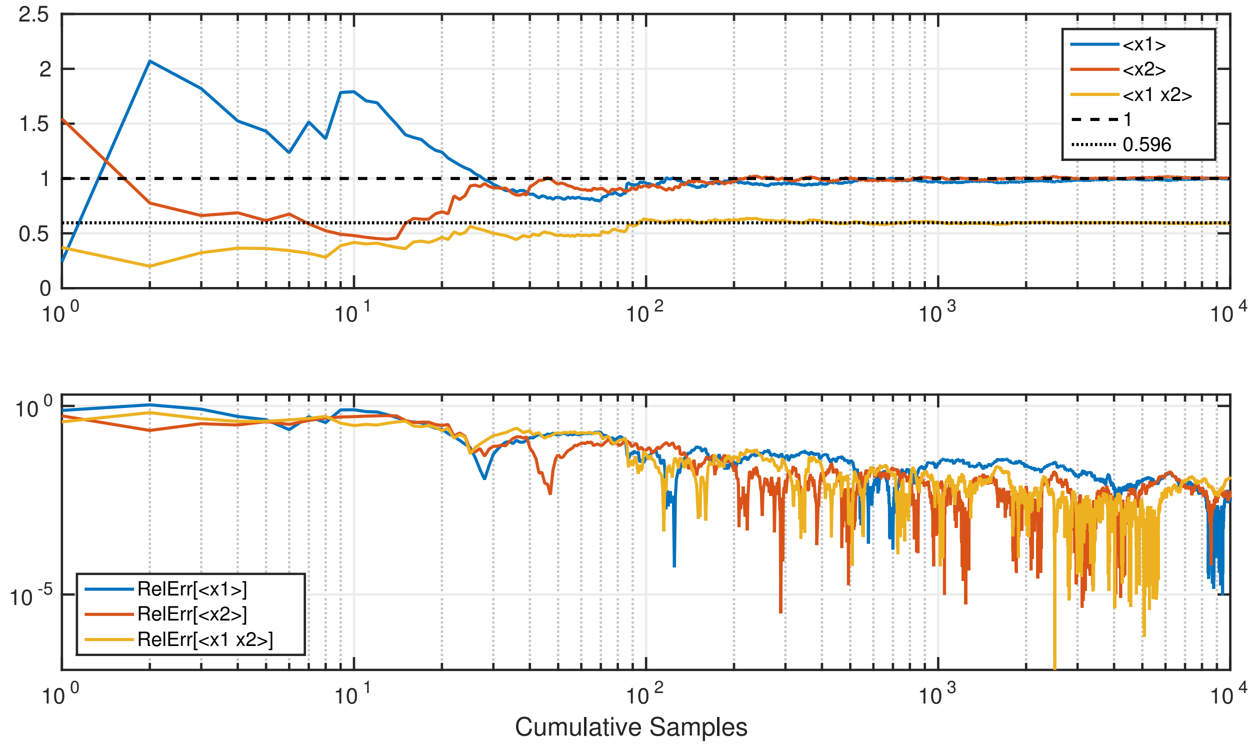

Figure 1: Expectation values of various functions of x_1 and x_2, using the first n cumulative samples. Analytical expectation values plotted to show convergence, as well as relative error.

Problem 3

The thermal coefficient K of a 1D slab is characterized by a lognormal random process

$$K(x,\omega) = \exp(G(x,\omega)), \qquad x \in (0,1),$$

where $G(x,\omega)$ is a Gaussian random process defined on $(0,1)$. The mean and covariance functions of the Gaussian process $G(x,\omega)$ are

$$\langle G(x) \rangle = 1.0, \qquad x \in (0,1),$$

and

$$C_{GG}(x_1,x_2) = \sigma^2 \exp\left(\frac{-|x_1 - x_2|}{\ell}\right), \qquad (x1,x2) \in (0,1) \times (0,1),$$

respectively. We would like to compute the statistics of the temperature field $u(x)$ by solving the governing steady-sate stochastic heat equation

$$\frac{\partial}{\partial x}\left(K(x,\omega)\frac{\partial u(x,\omega)}{\partial x}\right) = 1.0, \qquad x \in (0,1), \tag{6}$$

$$u(0,\omega) = 0, \tag{7}$$

$$u(1,\omega) = \theta(\omega), \tag{8}$$

where $\theta(\omega) \sim N(0,0.1)$ characterizes the uncertainty in the right boundary condition and is statistically independent from $K(x,\omega)$. Set $\sigma = 2$ and $\ell = 0.2$, and use the codes form Homework 1 to generate samples of $G_d(x,\omega)$ when only $d = 10$ terms in the KL expansion of $G(x,\omega)$ are used.

We write a standard second-order central-difference implicit code to solve the PDE in (8) for fixed ω, meaning that $K(x,\omega) \to K(x)$ and $\theta(\omega) \to \theta$. Next, we compute the mean $\langle u(x) \rangle$ and variance $\mathsf{Var}(u(x))$ of the solution using a Monte Carlo simulation, and verify that these statistics converge as the number of samples is increased. Finally, letting $u_{\max} = \max_x u(x)$, we compute the probability that the maximum temperature on the slab exceeds a certain threshold,

$$P\left(u_{\max} > \langle u_{\max} \rangle + 3 \cdot \sqrt{\mathsf{Var}(u_{\max})}\right). \tag{9}$$

Using the analytical eigensystem solution for the Karhunen-Loeve expansion, realizations of G, K, and u are shown in Figure 2. A spatial discretization is chosen that uses 1001 uniformly-spaced nodes on the interval $[0,1]$. As the number of realizations n is increased from 1 to $N = 100{,}000$, the mean and variance converge to the functions presented in Figure 3. Both boundary conditions are satisfied, since $u(0) = 0$ is fixed and $u(1) = \theta(\omega) \sim N(0, 0.1)$ matches in both mean and variance plots. Finally, we see the probability converge to a value near 0.016 in Figure 4. Though not shown in this figure, the final statistics pertaining to $u_{\max}$ obtained from 250,000 realizations are shown in Table 1.

Quantity	Value
$\langle u_{\max} \rangle$	0.12568
$\mathsf{Var}(u_{\max})$	0.03379
$P(\cdot)$	1.59 %

Table 1: Relevant quantities in (9) computed from 250,000 realizations using 1,001 grid points.

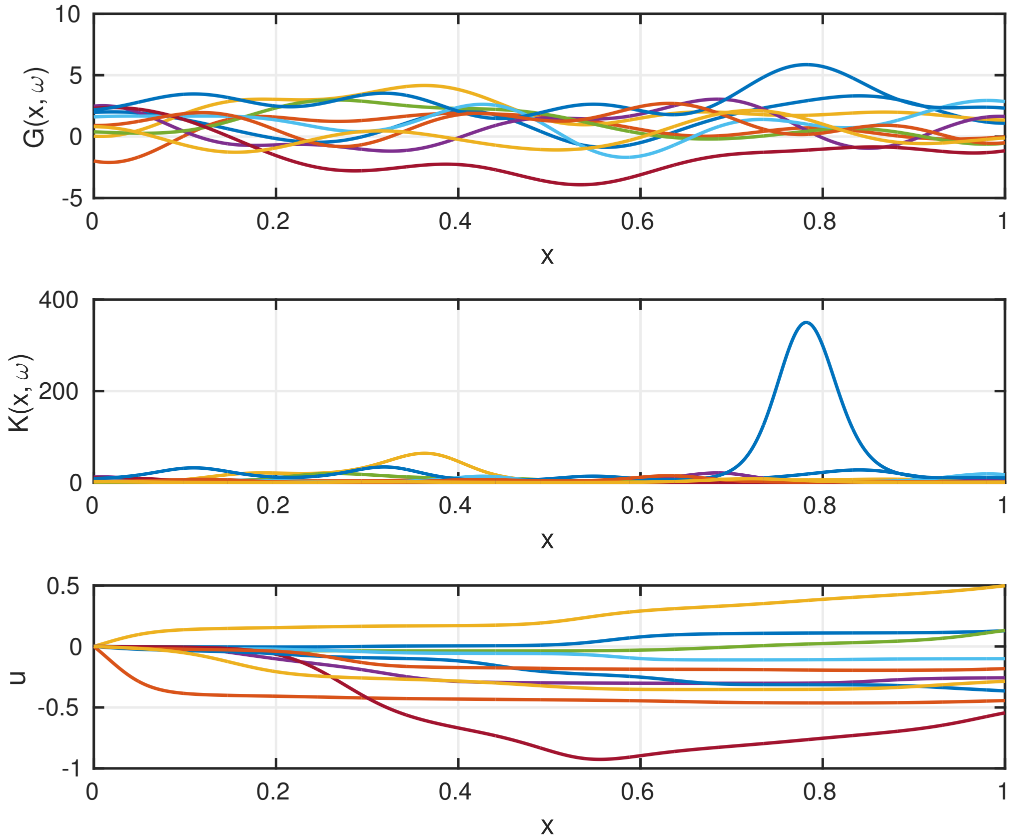

Figure 2: Ten realizations of G, K, and u.

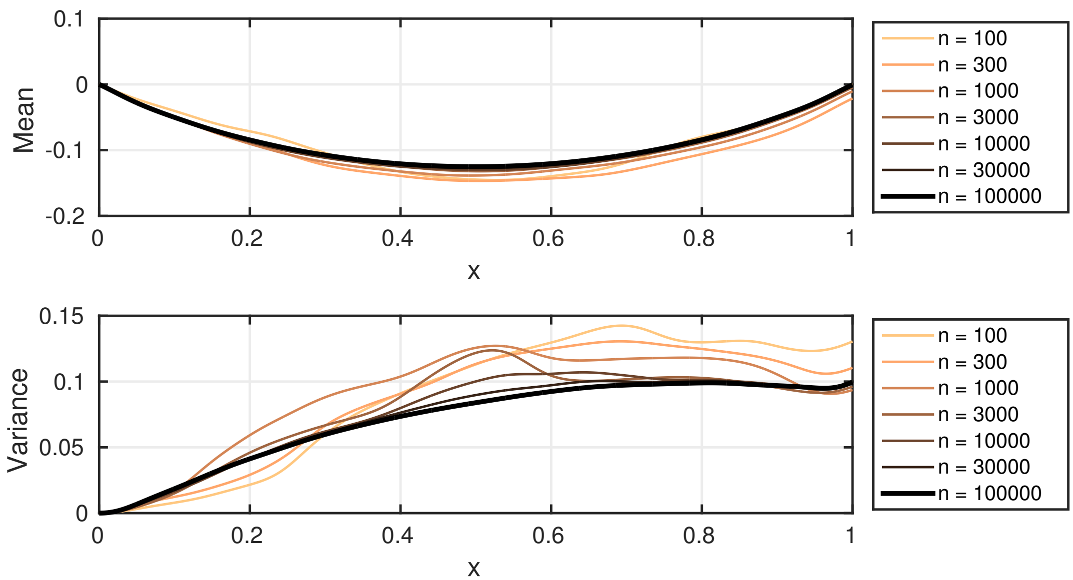

Figure 3: Convergence of the mean and variance of $u(x)$ as the number n of realizations is increased.

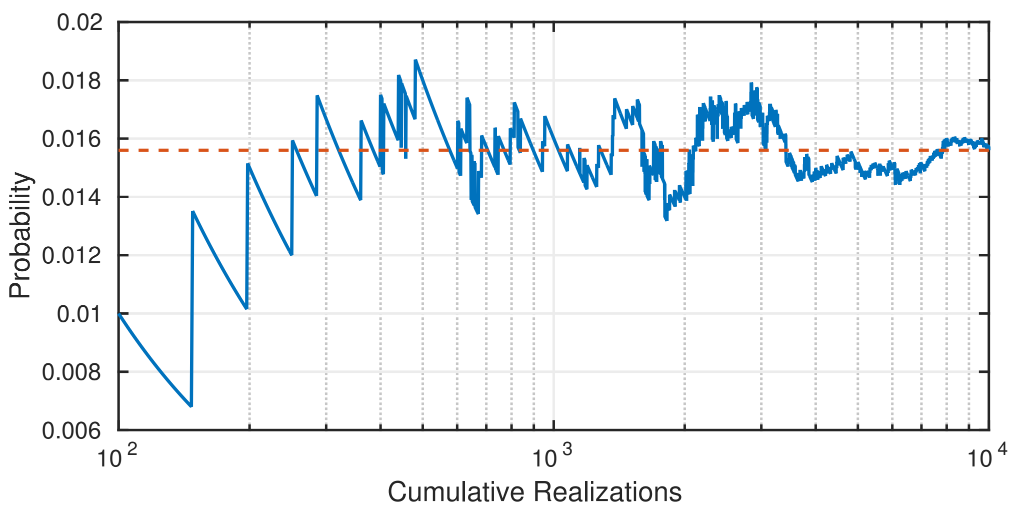

Figure 4: Convergence in the probability (9) as more realizations are generated. Final statistics use 2.5×10^5 realizations, but only up to 10^4 are shown here.