Supplemental Materials: Missing Proofs

Proof of Theorem 2. Part 1.

Note that (6) is equivalent to

$$P_\mu(P_{\xi|\mu}(f(x,\xi) \notin \mathcal{A}) \geq \alpha) \leq \beta \tag{17}$$

By the Markov inequality,

$$P_\mu(P_{\xi|\mu}(f(x,\xi) \notin \mathcal{A}) \geq \alpha) \leq \frac{E_\mu[P_{\xi|\mu}(f(x,\xi) \notin \mathcal{A})]}{\alpha}$$

So

$$\frac{E_\mu[P_{\xi|\mu}(f(x,\xi) \notin \mathcal{A})]}{\alpha} \leq \beta \tag{18}$$

guarantees (17). Note that (18) is equivalent to

$$E_\mu[P_{\xi|\mu}(f(x,\xi) \in \mathcal{A})] \geq 1 - \alpha\beta$$

or

$$P_{\mu,\xi}(f(x,\xi) \in \mathcal{A}) \geq 1 - \alpha\beta \tag{19}$$

where $P_{\mu,\xi}$ denotes the joint probability with respect to the posterior distribution of μ and the stochasticity of ξ given μ. The Monte Carlo scheme in PCSG precisely generates samples according to $P_{\mu,\xi}$. Theorem 1 implies that choosing N satisfying (8) in the sampled program (7) to obtain x guarantees (19), and consequently (6), with probability at least $1-\delta$. This concludes the proof of Part 1.

Part 2. Let $\mathcal{M}$ be the event

$$P_\mu(P_{\xi|\mu}(f(x;\xi) \in \mathcal{A}) \geq 1-\alpha) \geq 1-\beta$$

in the sample space under the Monte Carlo sample generation (that obtains x). Let $I_{\mathcal{M}}$ be the indicator function on the occurrence of $\mathcal{M}$. Then we have

$$\begin{aligned} & E_{MC}[I_{\mathcal{M}}] \\ = \;& P_{MC}(P_\mu(P_{\xi|\mu}(f(x;\xi) \in \mathcal{A}) \geq 1-\alpha) \geq 1-\beta) \\ \geq \;& 1-\delta \end{aligned} \tag{20}$$

by (9). Now, consider

$$\begin{aligned} & E_{MC}[P_\mu(P_{\xi|\mu}(f(x;\xi) \in \mathcal{A}) \geq 1-\alpha)] \\ \geq \;& E_{MC}[P_\mu(P_{\xi|\mu}(f(x;\xi) \in \mathcal{A}) \geq 1-\alpha) I_{\mathcal{M}}] \\ \geq \;& E_{MC}[(1-\beta) I_{\mathcal{M}}] \\ \geq \;& (1-\beta)(1-\delta) \end{aligned}$$

by using (20) in the last inequality. This concludes the proof.

□

Proof of Theorem 3. From Theorem 2 part 2, if we use β_t, δ_t and N_t satisfying (12), we have

$$\begin{aligned} & E_{MC}[P_{\mu_t}(P_{\xi_t|\mu_t}(f(x;\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) < 1-\alpha)] \\ \leq \;& 1-(1-\beta_t)(1-\delta_t) \end{aligned} \tag{21}$$

We want to show (14), which is equivalent to

$$\begin{aligned} & P_{MC}(P_{\mu_{1:T}}(P_{\xi_t|\mu_t}(f_t(x_t,\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) < 1-\alpha \\ & \text{for some } t \in \mathcal{S}) > \beta) \leq \lambda \end{aligned} \tag{22}$$

Note that the left hand side of (22) is bounded from above as

$$\begin{aligned} & P_{MC}(P_{\mu_{1:T}}(P_{\xi_t|\mu_t}(f_t(x_t,\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) < 1-\alpha \\ & \text{for some } t \in \mathcal{S}) > \beta) \\ \leq \;& P_{MC}\left(\sum_{t\in\mathcal{S}} P_{\mu_{1:T}}(P_{\xi_t|\mu_t}(f_t(x_t,\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) \right. \\ & \left. < 1-\alpha) > \beta \right) \\ \leq \;& \frac{1}{\beta} E_{MC}\left[\sum_{t\in\mathcal{S}} P_{\mu_{1:T}}(P_{\xi_t|\mu_t}(f_t(x_t,\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) \right. \\ & \left. < 1-\alpha) \right] \text{ (by the Markov inequality)} \\ = \;& \frac{1}{\beta} \sum_{t\in\mathcal{S}} E_{MC}[P_{\mu_t}(P_{\xi_t|\mu_t}(f_t(x_t,\xi_t) \in \mathcal{A}_t | \mathcal{F}_{t-1}) \\ & < 1-\alpha)] \\ \leq \;& \frac{1}{\beta} \sum_{t\in\mathcal{S}} (1-(1-\beta_t)(1-\delta_t)) \text{ (by using (21))} \end{aligned}$$

Hence

$$\frac{1}{\beta} \sum_{t\in\mathcal{S}} (1-(1-\beta_t)(1-\delta_t)) \leq \lambda \tag{23}$$

guarantees that (22) holds. Noting that (23) is equivalent to (13), we conclude our theorem. □

Proof of Proposition 1. Let $s = |\mathcal{S}|$. Setting $\beta_t = \delta_t = \gamma$, we have $\sum_{t\in\mathcal{S}}(\beta_t + \delta_t - \beta_t\delta_t) = s(2\gamma - \gamma^2)$. We want $s(2\gamma-\gamma^2) \leq \beta\lambda$, or equivalently

$$\gamma^2 - 2\gamma + \beta\lambda/s \geq 0 \tag{24}$$

Since the left hand side of (24) is a convex quadratic function, (24) holds if and only if

$$\gamma \geq 1 + \sqrt{1-\beta\lambda/s} \quad \text{or} \quad \gamma \leq 1 - \sqrt{1-\beta\lambda/s}$$

The first condition is never satisfied since γ must be ≤ 1. The second condition is valid and gives (14). □

Proof of Proposition 2. Without loss of generality, we label t as the counter in $\mathcal{S}$ for convenience (i.e., assume that $t = \zeta(t)$ by relabeling β_t and δ_t). We want $\sum_{t=1}^{\infty}(\beta_t + \delta_t - \beta_t\delta_t) \leq \beta\lambda$ holds so that (14) holds regardless of T. Note that

$$\sum_{t=1}^{\infty}(\beta_t + \delta_t - \beta_t\delta_t) = 2\sum_{t=1}^{\infty}\gamma_t - \sum_{t=1}^{\infty}\gamma_t^2 \tag{25}$$

We analyze the two terms of the right hand side of (25). By definition $\gamma_t = \eta$ if and only if $t \leq 1/\eta^{1/\rho}$. Thus, for the first term, we have

$$\begin{aligned}
2\sum_{t=1}^{\infty}\gamma_t &= 2\left(\lfloor\frac{1}{\eta^{1/\rho}}\rfloor \cdot \eta + \sum_{t=\lfloor 1/\eta^{1/\rho}\rfloor+1}^{\infty}\frac{1}{t^{\rho}}\right) \\
&\leq 2\left(\eta^{1-1/\rho} + \int_{\lfloor 1/\eta^{1/\rho}\rfloor}^{\infty}\frac{1}{u^{\rho}}du\right) \\
&\leq 2\left(\eta^{1-1/\rho} + \frac{1}{\rho-1}\frac{1}{(1/\eta^{1/\rho}-1)^{\rho-1}}\right)
\end{aligned} \tag{26}$$

For the second term in (25), we have

$$\begin{aligned}
\sum_{t=1}^{\infty}\gamma_t^2 &= \lfloor\frac{1}{\eta^{1/\rho}}\rfloor \cdot \eta^2 + \sum_{t=\lfloor 1/\eta^{1/\rho}\rfloor+1}^{\infty}\frac{1}{t^{2\rho}} \\
&\geq \frac{\eta^2}{\eta^{1/\rho}+1} + \int_{\lfloor 1/\eta^{1/\rho}\rfloor+1}^{\infty}\frac{1}{u^{2\rho}}du \\
&\geq \frac{\eta^2}{\eta^{1/\rho}+1} + \frac{1}{2\rho-1}\frac{1}{(1/\eta^{1/\rho}+1)^{2\rho-1}}
\end{aligned} \tag{27}$$

Therefore, combining (26) and (27) into (25), we have (15) implies $2\sum_{t=1}^{\infty}\gamma_t - \sum_{t=1}^{\infty}\gamma_t^2 \leq \beta\lambda$ or (14). □

Proof of Corollary 4. The corollary follows immediately by noticing that the linear program (16) is trivially convex, and applying Theorem 3. □